Assembly and Structure of Bacteriophage Capsids and Tails

by

Patricia L Campbell

B.S., Columbus State University, 2011

Submitted to the Graduate Faculty of

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Patricia L Campbell

It was defended on

July 25, 2019

and approved by

William Furey, Professor, Department of Pharmacology and Chemical Biology

Fred Homa, Professor, Department of Microbiology and Molecular Genetics

Jeff Brodsky, Professor, Department of Biological Sciences

Dissertation Director: Robert Duda, Department of Biological Sciences

Dissertation Director: Roger Hendrix, Department of Biological Sciences

Dissertation Director: James Conway, Department of Structural Biology

Assembly and Structure of Bacteriophage Capsids and Tails

Patricia L Campbell, PhD

University of Pittsburgh, 2019

Phages HK97 and λ are model systems that have aided in the understanding of capsid and tail assembly. HK97 capsid assembly involves the formation of a T=7 icosahedral lattice from 415 monomers of the major capsid protein (mcp), 12 monomers of portal protein, and ~120 copies of the protease. Salt bridges have been found to play a critical role in the assembly pathway, but it is not clear how capsid size is determined. Phage phi1026b is larger than HK97 but the mcp shares 50% sequence similarity with HK97. We hypothesize that comparing the atomic models of phi1026b and HK97 will help illuminate conserved and unique interactions that may play a role in assembly and potentially size determination. By utilizing the FEI Krios microscope, Falcon II camera, and EPU data acquisition software, high-resolution structures were generated which allowed for the generation of atomic models of the phi1026b Prohead I and Head capsids. Differences were observed in the organization of the delta domain, capsomers, and intra-capsomer interactions. The E153-R210 salt bridge that is maintained throughout HK97 capsid assembly is not present between two monomers in the phi1206b Prohead I structure. Phage λ tail assembly proceeds in a stepwise manner in which tail proteins form an initiator complex onto which the major tail tube protein, gpV, polymerizes and is terminated by the tail termination protein, gpU. The tube domain of gpV has been studied in its monomeric form using NMR but the in vivo form has not been visualized. By generating a subnanometer structure for the λ tail tube, we are able to show how gpV undergoes structural changes to facilitate assembly. Also, a putative initiator complex comprised of the tail proteins gpG, gpT, gpH, gpM, gpL, gpI, and gpJ was isolated and

characterized using microscopy. Finally, a subnanometer structure for the λ tail tip complex (TTC) is presented providing some initial insight into the organization of the λ TTC. Overall, this research intends to optimize structure determination with the cryo-EM, thereby facilitating studies on the structure and assembly of large viruses.

Table of Contents

List of Tables

List of Figures

Preface

Thank you, James and Roger, for giving me the opportunity to be a member of your labs and for sharing your great knowledge with me. Roger, it was a profound honor to be your student and you will be greatly missed. James, I deeply value and respect that you provided the chance for me to enter this grad program and shared the joys of electron microscopy with me. I will miss the daily coffee breaks, the great range of conversation topics ranging from politics to science, and the humor. Thank you, Bob and Alexis, for spending the time to both train me and have in depth conversations about phage, electron microscopy, reconstructions, and analysis. Reiko, thank you for your continued support and profound insight throughout the years. I would also like to thank Craig and Susan for their advice and guidance. Thank you, Sasha, for your insight and knowledge.

Thank you, Bonnie, Aletheia, Josh, Jake, and Jianfei for being so supportive and great teammates. Thank you, Ryan, Ben, Sean, and Abhishek for being such great classmates and friends.

Thank you to all the friends for your support and advice: Danielle, Lauren, Isela, Bonnie, Tamara, Jocelyn, Aletheia, Mike C, Caitlin, Ayne, Ciaran, Nilesh, Tessa, Karen, Remy, John G, Greg, James J, Chelsea, Nicole, Stephanie, Mindy M, Mindy C, Justin V, and so many more.

Thank you to my Mother, Father and Grandmother, Nance and Alex for your support, patience, and encouragement. I also want to thank my cat, Harold, for being a cat - sometimes, that’s exactly what’s needed.

1.0 Introduction

1.1 Bacteriophages and Structural Biology

The viruses that feed on bacteria are called bacteriophages or phages. There are an estimated 10^{31} phages in the world (Bergh *et al.*, 1989; Brüssow & Hendrix, 2002; Grose & Casjens, 2014; Hendrix, 2003) whose interactions with bacteria have the ability to affect the health of animals and ecosystems (Brussaard *et al.*, 2008; Tetz *et al.*, 2018). Bacteriophages were discovered independently by Twort and D'herelle when they observed an unknown agent that could kill bacteria (D'Herelle, 1917; Twort, 1925). The debate about the nature of these antimicrobial agents wasn't resolved until they were visualized using an electron microscope in 1940 by Herman Ruska whose brother is credited as having invented that technology (Ackermann, 2011; Luria *et al.*, 1943; Ruska, 1940). Ruska's images confirmed that the agent had a particulate nature, and electron microscopy has been used since to classify phages according to their morphology (Bradley, 1967). Other structural techniques such as NMR and Xray crystallography provided insight into potential evolutionary relationships among bacteriophage and other biological systems through identification of conserved morphology and protein folds (Krupovič & Bamford, 2008). The work presented in this thesis exploits recent advances in cryo-electron microscopy (cryo-EM) technology to study both the structure and assembly pathways of the capsids and tails of *Siphoviridae* bacteriophages.

1.2 The Morphological Diversity of *Caudoviridae*

As of 2012, the vast majority (96.3%) of the roughly 6,300 phages described belong to the taxonomic order *Caudoviridae* or the "tailed" phages which feature a linear dsDNA genome, an icosahedral capsid, and a tail (Ackermann & Prangishvili, 2012). Within *Caudoviridae*, there are three families that are defined by their tail morphology (Figure 1). The *Myoviridae* (myo "muscle") family have a long, stiff, contractile tail with two concentric tubes and a tail core surrounded by a contractile sheath. The *Podoviridae* (podo "foot") family have a short non-contractile stubby tail which contains a tail spike and needle. The *Siphoviridae* (sipho "straw") family have a long, flexible, non-contractile tail comprised of a single tube. Phages in all three families often have long tail fibers that help mediate host recognition and adsorption. The baseplate, found in both *Siphoviridae* and *Myoviridae*, is responsible for attaching to the host and triggering genome injection. Some *Siphoviridae* have a simpler and relatively smaller tail tip in place of the baseplate that performs the same functions. *Podoviridae* have no tail tube but possess an inner core in their capsid which is made of proteins required for host entry.

The capsid is the proteinaceous shell that all three families of *Caudoviridae* use to encapsulate their viral genomes. These shells have icosahedral symmetry although their sizes vary considerably and in some cases the head is elongated along one axis into a 'prolate icosahedral' capsid. While the capsid is made of a major capsid protein (mcp), a portal protein, and optional minor proteins, the assembly of the capsid requires the aid of a scaffold protein which is not present in virions. Some capsids have more than one capsid shell protein, including the T4 (*Myoviridae*) which has a major protein that form hexamers, gp23, and a minor protein that forms pentamers, gp24 (Rao & Black, 2010). Capsids may also have additional decoration proteins that bind at local 3-fold symmetry axes, as in gpD of phage λ (Yang *et al.*, 2000), or the at center of the hexamers,

as in pb10 of phage T5 (Vernhes *et al.*, 2017). The role of decoration proteins can vary widely with some having no known function, as in gp12 of phage SPP1 (Zairi *et al.*, 2014), to being critical to capsid stability and infection, as in Psu in phage P4 (Christie & Dokland, 2012). The portal is a dodecamer that sits at one vertex of the capsid, facilitates genome packaging and release, and is the site of tail attachment. *Podoviridae* have an inner capsid core composed of proteins required for host entry. This study focuses on the capsids and tails of *Siphoviridae*.

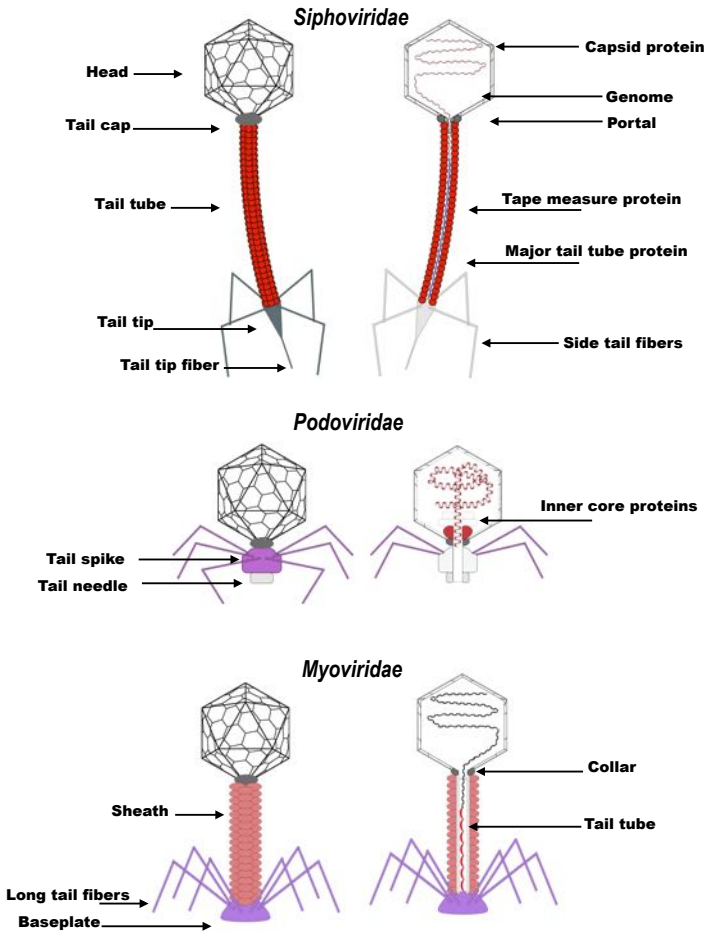

Figure 1. Schematics outlining Caudovirales morphology

All three families have an icosahedral head that encapsulates the genome and with a portal occupying one of the 12 vertices. *Siphoviridae* have long, non-contractile, flexible tails that connect to the tail tip complex (TTC). *Podoviridae* have a short tail spike and tail needle. *Myoviridae* have a long, contractile, rigid tail that connects to a baseplate.

1.3 Icosahedral Capsid Symmetry and Triangulation Numbers

Caudoviridae have icosahedral or prolate icosahedral capsids comprised of pentamers and hexamers of major capsid protein that assemble into closed shells. The icosahedron has 12 vertices and 20 equilateral triangular faces, and has 2, 3, and 5-fold rotational symmetries (Baker *et al.*, 1999). For genomic efficiency, the ideal way to construct such a seemingly complex structure is with many copies of a small subunit that can adopt multiple conformations as opposed to singular large pieces. To define how complex structures can be built with simple subunits, Caspar and Klug proposed that identical monomers can occupy quasi-equivalent environments (Caspar & Klug, 1962). The complexity of the icosahedral lattice is defined by a Triangulation number, $T = h^2 + hk + k^2$, where h and k are positive integer indices defined as "steps" in different directions between sites of local 6-fold symmetry that describe the shortest path from one vertex to the next. The T number also corresponds to the number of quasi-equivalent positions, so 60•T is the number of subunits in the capsid. The simplest capsid (T=1) is comprised of 60 monomers (12 pentamers) and each monomer is in an equivalent conformation and environment. Prolate icosahedral capsids have an elongated cylindrical section with icosahedral end-caps that are described by the T number and the lattice of the cylindrical section is described by a Q number (Baschong *et al.*, 1988; Moody, 1999).

1.4 Tail Structure

The three *Caudoviridae* families have a tail structure attached to portal-containing vertex on the capsid. *Podoviridae* have a short stumpy tail and a tail needle. *Siphoviridae* and *Myoviridae*

have several common features: a tail tube that the genome exits through, a baseplate/tail tip complex at the base of the tail, and one or more fibers. At the top of the tail tube is the 'collar' or 'neck' which is composed of the head-tail connector proteins, the tail terminator protein, and in some cases 'whiskers' which are fibrous proteins that aid in environment sensing. As the name suggests, the tail tube is a cylinder with a ~3 nm inner diameter just wide enough for a dsDNA strand to pass through. In *Siphoviridae* and *Myoviridae,* the tube is made from of a single protein, the tail tube protein, that forms multimeric rings stacked on top of each other. Tail tubes generally have 6-fold helical symmetry although pseudo-hexameric symmetry has been found in phiCbK (Leonard *et al.*, 1973) and T5 (Arnaud *et al.*, 2017)(both *Siphoviridae*). For *Myoviridae*, an additional cylinder called the sheath surrounds the tail tube and also features 6-fold helical symmetry. Inside *Siphoviridae* and *Myoviridae* tails is an elongated tape measure protein (tmp) that determines the correct length of the tail and plays a role in host entry (Abuladze *et al.*, 1994; Katsura & Hendrix, 1984).

The three phage families have different processes for introducing their dsDNA into their hosts. The *Podoviridae* family display a range of cell entry strategies though some of the best studied are those of phages P22 and T7. When the *Podoviridae* tail spike contacts the host, proteins in the tail cleave the peptidoglycan wall allowing proteins located inside the capsid to pass through the portal vertex, cross the periplasmic space, and reorganize in the host membrane to generate a conduit for the genome to pass through (Figure 2E, F) (Casjens & Molineux, 2012). The *Myoviridae* family features a baseplate that makes contact with the host, triggering a conformational change that contracts the tail sheath and pushes the inner tail tube into the host to allow passage of the genome into the host (Figure 2C, D) (Leiman & Shneider, 2012). Information about *Siphoviridae* host entry is limited compared to *Myoviridae*. It appears that the *Siphoviridae*

receptor binding protein makes contact with host and may be cleaved or dissociated to enable the tmp protein to form a channel that allows the passage of the genome into the host (Figure 2A, B) (Boulanger *et al.*, 2008; Davidson *et al.*, 2012). It has been proposed that *Siphoviridae* genome release is triggered by a dominos-like rearrangement of tail tube proteins from the tail tip to the portal (Plisson *et al.*, 2007) or the loss of the tmp from the tail tube (Arnaud *et al.*, 2017; Konopa & Taylor, 1979; Scandella & Arber, 1976).

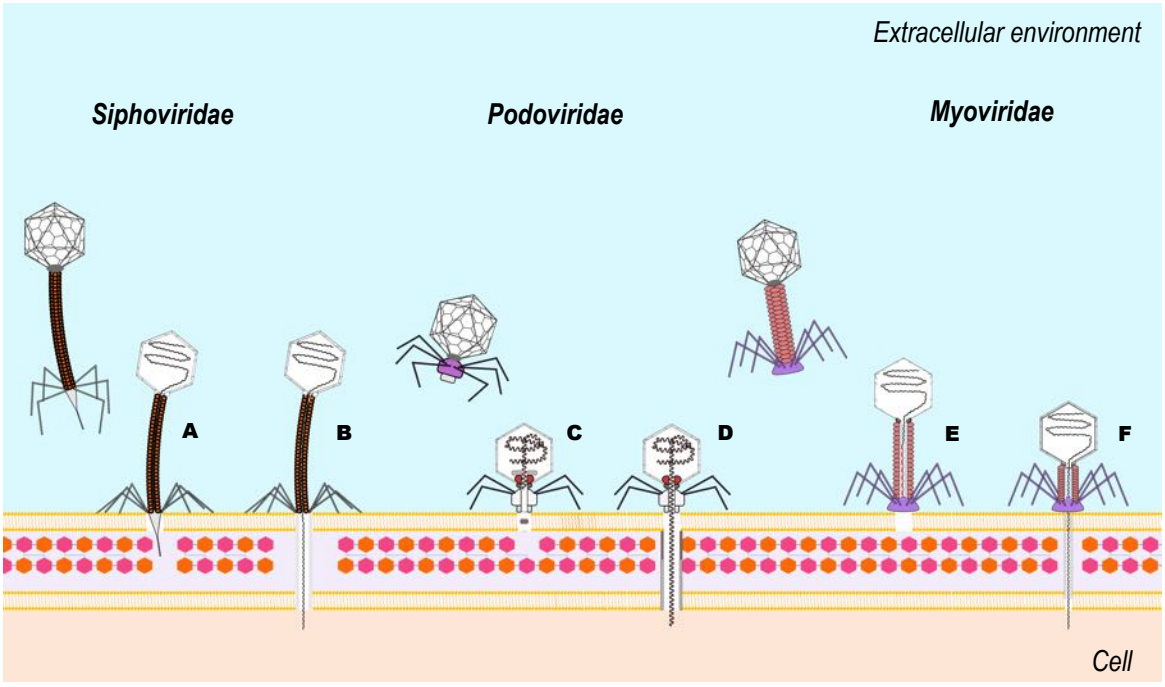

Figure 2. Host entry strategies of Caudovirales infecting Gram-negative bacteria

A) The *Siphoviridae* TTC makes contact with the host cell which then cleaves the peptidoglycan cell wall. B) The tmp protein and host proteins form a conduit for the phage genome to pass through into the host. C) The *Podoviridae* tail needle makes contact with the host cell and proteins in the tail spike cleave the peptidoglycan cell wall. D) Proteins in the *Podoviridae* capsid pass through the portal and play a role in forming a conduit for the phage genome to pass through into the host. E) The *Myoviridae* baseplate and tail fibers make contact with the host cell. F) Proteins in the *Myoviridae* baseplate go through a conformational change that triggers the sheath to contract, forming a conduit for the phage genome to pass through.

1.5 Tail Tips and Baseplates

Tail tip complexes (TTCs) are located at the distal end of the tail and play critical roles both in assembly, host identification, and genome injection. *Siphoviridae* TTCs can either be bulky hexagonal structures, called baseplates, or simpler, elongated structures called tail tips. *Myoviridae* TTCs are usually just referred to as baseplates. Conserved features of the *Siphoviridae* TTCs include the distal tip protein (DTP or Dit), the receptor binding protein (RBP), the baseplate hub protein (BHP, which sits at the center of the tail tip), and central fiber protein (CFP). In most phages there are long side fiber proteins (SFP) that are connected to the sides of the tail tip. The proteins of Gram-positive-infecting *Siphoviridae* are better characterized than those of Gram-negative-infecting *Siphoviridae*. The DTP can act as a connector between the tail tip and the tail tube. pb9, the DTP protein of bacteriophage T5 (*Siphoviridae)*, one of the few Gram-negative-feeding *Siphoviridae* TTC proteins to be structurally characterized, shares structural similarity with Gram-positive feeding *Siphoviridae* Dit proteins (Flayhan *et al.*, 2014). In bacteriophage T4 (*Myoviridae)*, the BHP generally joins the 6-fold symmetric tail tube to the 3-fold symmetric tail spike (Kostyuchenko *et al.*, 2003; Leiman *et al.*, 2010). Structural homology between the BHP of bacteriophage Mu (*Myoviridae)* and Gram-positive-feeding *L. lactis* bacteriophage p2 (*Siphoviridae)* suggests conservation in the tail proteins between the *Caudoviridae* families or convergent evolution (Sciara *et al.*, 2010). The BHP gp18 of the Listeria phage A118 (*Siphoviridae*) shares structural homology with many Gram-positive-infecting *Siphoviridae* (Davidson *et al.*, 2012; Kondou *et al.*, 2005). While there are currently no atomic models of Gram-negative bacteria feeding *Siphoviridae* BHPs, structural predictions suggest possible structural similarity with currently available BHP atomic models (Zivanovic *et al.*, 2014). The central fiber

protein along with the receptor binding protein makes contact with the host cell and is referenced as the tail 'spike' or 'fiber'.

1.6 Assembly Pathways

Bacteriophages usually have consistent capsid and tail sizes suggesting that there are elements that regulate size with considerable fidelity. Phage assembly proceeds in a stepwise manner with each step signaling the following step. Early discovery of these pathways came from studies of amber mutants of structural genes that generated incomplete or aberrant particles in both head and tail assembly (Casjens & King, 1975; Edgar & Lielausis, 1968). In many cases, conformational changes in structural proteins control assembly by preventing the aberrant addition of further structural proteins or premature conformation changes in structural proteins. During tail assembly, the tail tube cannot form prior to the formation of a stable baseplate or tail tip complex (King, 1968). In capsid assembly pathways, steps proceed with both large conformational changes such as capsid expansion and small conformational changes such as local changes in the domains of the mcp (Conway *et al.*, 1995). For *Myoviridae* and *Siphoviridae*, tail and head assembly occur along two separate pathways with the final step of assembly being the joining of the tail and the head to form the mature virion (Casjens & Hendrix, 1988).

Assembly usually produces capsids of consistent size (Figure 3C). The major capsid protein, with the aid of a scaffolding protein, forms a closed, spherical immature procapsid made of hexamers and pentamers with the portal occupying one vertex (Aksyuk & Rossmann, 2011; Casjens & Hendrix, 1988; Suhanovsky & Teschke, 2015). Once the procapsid is assembled, the scaffolding proteins are removed from the structure by diffusion or proteolysis resulting in an

empty shell. A phage encoded protease is co-assembled into the procapsids of viruses that use proteolysis for scaffold removal. The genome is packaged through the portal and the capsid expands to produce the angular mature head. After packaging and expansion, some viruses undergo further head stabilization by the addition of decoration proteins, such as gpD in λ (Yang *et al.*, 2000), or covalent modification, such as the crosslink between Lys169 and Asn356 in HK97, (Duda, Hempel*, et al.*, 1995).

Tail assembly follows a pathway in which multiple structural proteins bind in highly regulated steps. In *Podoviridae*, the tail proteins are added to the completed head in a stepwise manner (Figure 3E). In *Myoviridae* and *Siphoviridae*, tail assembly proceeds separately from head assembly (Figure 3A, B). The first stage of tail assembly is the formation of the baseplate or TTC. The T4 (*Myoviridae*) baseplate assembly involves 130 subunits of 14 separate proteins and is organized as 6 'wedges' that surround a 3-fold symmetric 'hub' (Kikuchi & King, 1975; Leiman *et al.*, 2010; Yap *et al.*, 2016). Once complete, the baseplate or TTC forms the tail assembly initiator complex by recruiting the tape measure protein (tmp) that regulates the length of the tail. The initiator complex serves as the site of the tail tube protein polymerization which occurs along the length of the tmp until the correct length is reached (Figure 3B) (Abuladze *et al.*, 1994; Katsura & Hendrix, 1984; Xu *et al.*, 2014). *Myoviridae* tail assembly involves an additional assembly step after the formation of the tail tube in which the sheath protein polymerizes around the completed tail tube (Figure 3A). Tail tube polymerization is terminated by a tail termination protein and then the tail is "capped" (Katsura & Kuhl, 1975; Vianelli *et al.*, 2000). The fully assembled tail is then attached to a DNA-filled completed head at the portal to yield a functional mature virion.

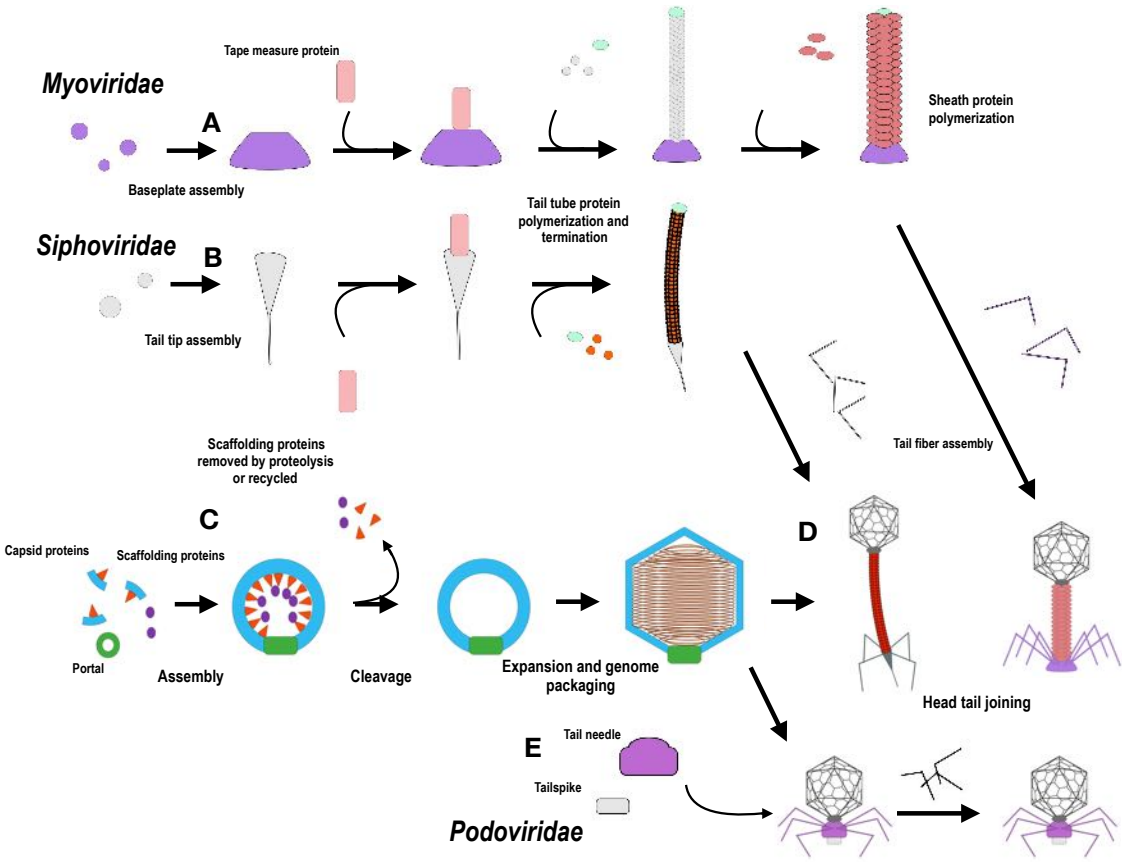

Figure 3. Assembly pathways for the *Caudoviridae* families

A) The *Myoviridae* tail assembly starts with the formation of the baseplate. Once finished, the tmp connects with the baseplate followed by tail tube polymerization along the length of the tmp. Once tail tube polymerization is complete, the sheath protein polymerizes around the tail tube. B) *Siphoviridae* tail assembly also begins with the assembly of the tail tip or baseplate, depending on the phage. Once the tail tip and tmp have formed the initiator complex, the tail tube protein polymerizes along the length of the tmp and is terminated by the tail termination protein. C) Capsid assembly for the three families proceeds as outlined in this step. The mcp, portal, scaffolding protein and any additional proteins such as protease assemble into the immature form of the capsid, the prohead. The scaffolding proteins are either removed by proteolysis or diffusion. DNA is packaged through the portal triggering expansion resulting in the mature capsid, the head. D) For *Myoviridae* and *Siphoviridae*, once tail and capsid assembly are complete, the tail is added to the portal-containing vertex of the head to produce the mature virion. E) *Podoviridae* tail assembly involves the addition of the tail proteins directly to the portal-containing vertex on the head.

1.7 Models of Study

This work will focus on the tail of λ phage and the capsid of phi1026b. λ phage tail is a model system that the Hendrix/Duda lab has studied for many decades and has served as one of

the primary model systems for *Siphoviridae* tail assembly and structure. In addition, λ has played a critical role in genetics, gene expression, and biotechnology such as phage display (Murray & Gann, 2007; Nicastro *et al.*, 2014). Phage phi1026b shares 50% sequence identity with the mature major capsid protein of HK97, which the Hendrix/Duda and Conway labs have used as a model system for studying the assembly and structure of capsids, including how a single protein can regulate assembly and size determination of the capsid.

1.7.1 Phage λ Structure

Phage λ is ~230 nm in length with an icosahedral head, a flexible noncontractile tail with a conical tip, and a single tail fiber. The λ head has T=7 geometry and is comprised of the major capsid protein E with the capsid decoration protein D attached as a trimer at the sites of 3-fold symmetry (Dokland & Murialdo, 1993; Yang *et al.*, 2000) (Figure 5). The capsid is 60 nm in diameter and encapsulates the 48,502 bp double stranded DNA phage genome. The portal complex, a dodecamer of the portal protein gpB*, occupies one of the 12 vertices and is connected to the head-tail joining protein W and tail attachment protein FII that form the junction between the head and the tail. The tail is comprised of a 135 nm long tail tube, 15 nm long conical tip, and a 23 nm long tail fiber (Katsura, 1983). The tail terminator protein gpU 'caps' the tail and connects it to the protein complex at the portal containing vertex. The tail tube has 32 hexameric disks of the major tail tube protein gpV that enclose the tape measure protein H. The conical tip at the tail base may contain some or all of the proteins gpL, gpM, gpI, gpK. At the base of the conical tip is the tail fiber protein J that recognizes the λ receptor protein, LamB, on the *E. coli* host. Phage Ur-λ, the original isolate of phage λ, has 4 side tail fibers attached to the side of the conical tip that are not present in the widely studied λ PaPa (Hendrix & Duda, 1992). This is because λ PaPa

features a frameshift mutation in gene *stf* that results in two open reading frames, ORF 401 and ORF 314, where gpORF 401 may comprise the barely visible tail fibers of λ PaPa (Haggård-Ljungquist *et al.*, 1992).

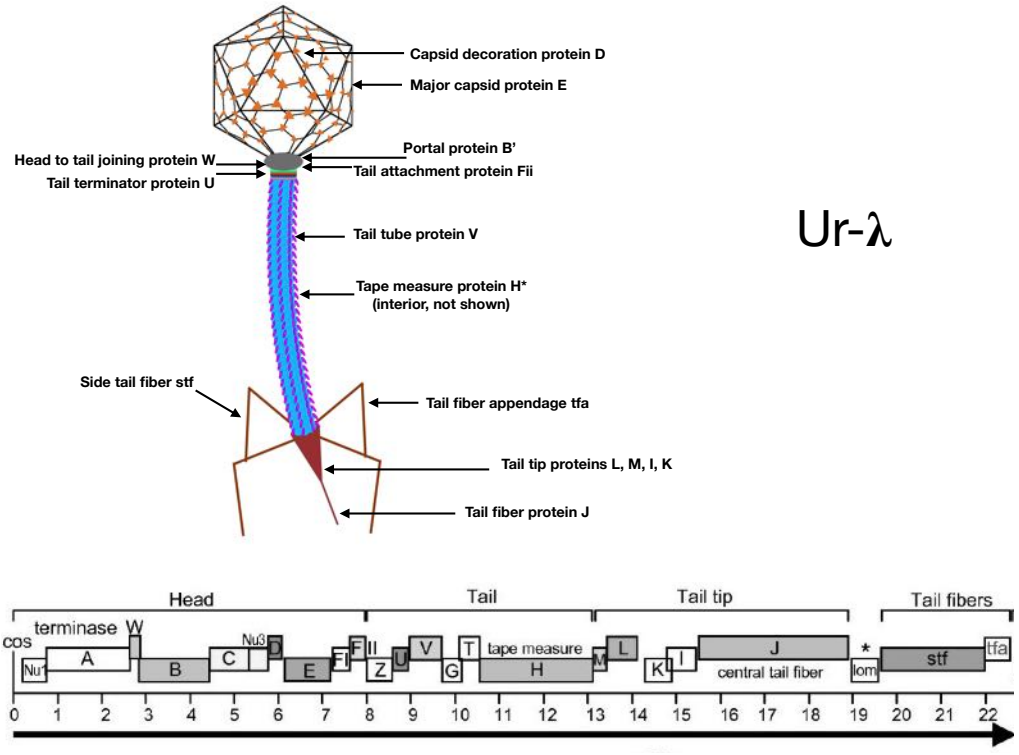

Figure 4. Anatomy of Ur- λ and structural genes

Image was adapted and altered from Rajagopala 2011 (Rajagopala *et al.*, 2011)

1.7.2 Phage λ Tail Assembly

Phage λ tail assembly requires 11 genes and proceeds in a stepwise manner. Overall, 21 λ structural proteins are required for the assembly of the mature λ virion. Phage genes Nu1, A, W, B, C, Nu3, D, E, F_I, F_{II} are required for head assembly, while genes Z, U, V, G, T, H, M, L, K, I, J are required for tail assembly (Figure 5). The genes *stf* and *tfa* are required to form the side tail

fibers but they are not required to generate viable tails (Xu, 2001). The tail genes are expressed in a single transcript but translation levels of each are generally comparable to the amount found in mature tails. Exceptions include gpU, gpV and gpG which are expressed in higher concentrations in vivo relative to other structural proteins. λ tail assembly is divided into 3 major steps: the formation of the initiator, the polymerization of the major tail protein, and the termination of major tail protein polymerization (Figure 6) (Katsura, 1990).

The initiator is comprised of the proteins gpJ, gpI, gpL, gpK, gpM, gpH, gpG and gpGT (Katsura & Kühl, 1975)(Figure 6C). The TTC proteins, gpJ, gpI, gpL, gpK, associate independently of gpG, gpGT, gpH and gpM. TTC assembly starts with the 15S particle, which is comprised of 3 copies of gpJ, and continues with the addition of gpI, gpL, and gpK proteins to form the $15S_{[II]}$ complex (Figure 6A). Separately, the chaperone proteins, gpG and gpGT, associate with gpH, the tmp (Xu *et al.*, 2013). The gpG/gpGT/gpH complex binds to the $15S_{[II]}$ complex forming the $25S_{[I]}$ complex, which can be converted to the true and final initiator of tail tube polymerization, the $25S_{[II]}$ complex, by the addition of gpM.

Polymerization of gpV onto the $25S_{[II]}$ initiator particle proceeds quickly. The current model postulates that monomeric gpV binds to the T domain of gpGT and mediates gpV polymerization (Xu *et al.*, 2013, 2014)(Figure 6D). In solution, gpV exists in a monomer-dimer equilibrium (Katsura & Tsugita, 1977) suggesting that gpV can associate independently of the initiator but polymerization requires the presence of the initiator. A model for how gpV changes when polymerized is based on a comparison with a structurally similar protein, the T4 tail tube protein gp19. The monomeric gpV atomic model has a disordered loop that is ordered in the polymerized gp19 suggesting that this region becomes ordered during polymerization (Pell *et al.*,

2009). As gpV polymerizes, gpG and gpGT are displaced from the initiator complex and polymerization continues until the full length of the tmp is reached.

The final step of tail assembly occurs when gpV polymerization stalls after reaching the full length of gpH (Katsura, 1976b) and 6 copies of gpU add to the end of the tail tube to stop further polymerization. gpU was identified as the tail terminator protein (TrP) when it was observed that mutants in the gene resulted in polytails, which are inactive tails of irregular length (Mount 1968). Further, U amber mutants ultimately lose gpH after gpV polymerization stalls which eventually resumes and continues to polymerize into polytails (Katsura, 1976b; Weigle, 1966). After the addition of gpU, ~100 amino acids are cleaved from the C-terminus of gpH yielding gpH* (Murialdo & Siminovitch, 1972; Tsui & Hendrix, 1983; Walker *et al.*, 1982). The role this cleavage plays as well as the fate of the cleavage product is unknown. gpZ is required to 'activate' the tail (Figure 6E) to complete tail assembly. It has been proposed that gpZ plays a role in identifying the right terminus of the λ genome and inserting into the tail although this is still a matter of debate (Thomas *et al.*, 1978). Once tail assembly is finished, the tail can bind to the portal on the completed head.

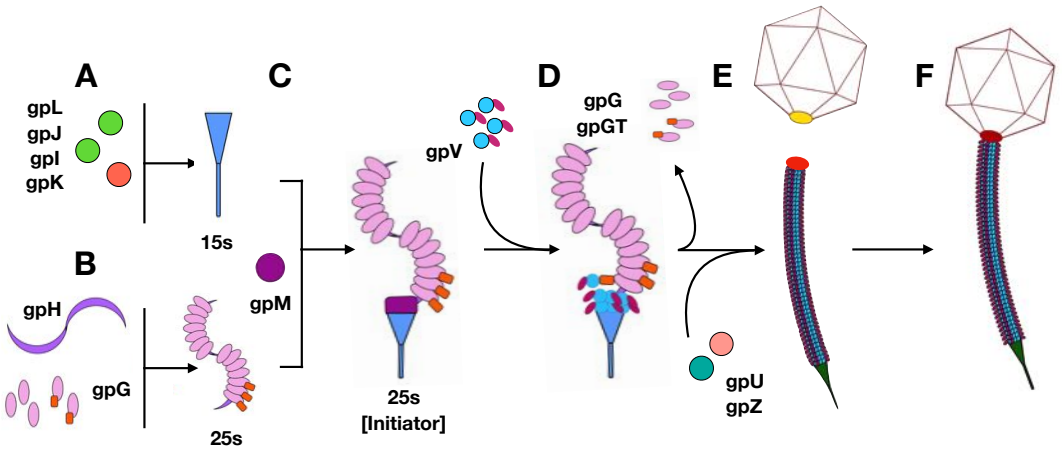

Figure 5. Assembly pathway of the λ tail tube

A) The formation of the 15S assembly intermediate by the tail tip proteins gpL, gpJ, gpI, and gpK. B) The associate of the tmp, gpH, with its chaperones gpG and gpGT. C) The formation of the 25S initiator complex by the proteins gpL, gpJ, gpI, gpK, gpH, gpG, gpGT, and gpM. D) The polymerization of the tail tube protein, gpV, on the initiator. E) The proteins gpU and gpZ stop polymerization and activate the tail for head attachment. F) Once the tail is activated, it is added to the portal containing vertex on the mature capsid.

1.7.3 Phage λ Tail Tube

The 135 nm long λ tail tube is primarily comprised of gpV, the major tail tube protein. Thirty-two hexameric rings of gpV form the shaft of the tail making it the most abundant protein in the tail. At pH levels below 2.8 and above 11.8, the tail dissociates into individual gpV hexamers with an interior empty space 3 nm in diameter (Bleviss & Easterbrook, 1971; Katsura, 1981). Analysis of these hexamers showed an inner ring of density that is 9 nm in diameter with outer protrusions making the total structure 18 nm in diameter (Katsura, 1981). In phage particles treated with protease, gpV is cleaved into two fragments (8 kDa and 17 kDa in size), but only the larger fragment remains associated with the virion (Roessner & Ihler, 1984). Viable deletions within gpV resulted in fragments as short as 18 kDa and still remained functional (Katsura, 1981). Together, these results show that gpV has two domains: a larger domain that forms the hexameric ring and a

smaller domain that forms external protrusions. Both domains have been studied using NMR spectroscopy. The tail tube domain of gpV is the N-terminal 157 residues of the 246 residue length protein (Pell *et al.*, 2009) and the smaller domain of gpV is formed by the C-terminal 89 residues that form 'protrusions' and feature a Ig-like fold (Pell *et al.*, 2010).

The tmp, gpH, is expected to occupy the interior of the tail tube and is the second most abundant component of the tail (Hendrix & Casjens, 1974). gpH serves three functions: formation of the initiator, measurement of the tail length, and DNA injection into host cells. gpH was identified as the tmp in a series of genetic experiments and observations that showed: 1) alterations around gene H resulted in tails of different length; 2) lambdoid phages with alternatively sized tails had corresponding differences in the gpH homologs; 3) H amber mutants generated tails of random length; 4) and the length of gpH polypeptide would directly correlate to the length of the tail if extended 'like a thread' (Katsura, 1976b; Katsura & Hendrix, 1984; Youderian, 1978). The majority of gpH is predicted to be primarily alpha helical with a glycine-rich C-terminus (Katsura, 1990). The C-terminus is required for the formation of the initiator and may also be the site of cleavage that reduces gpH to gpH* during assembly (Katsura, 1990; Murialdo & Siminovitch, 1972). Conversely, the majority of the N-terminus does not appear to be required for assembly and gpU binding but is present in the mature virion (Walker *et al.*, 1982). While 3-6 copies of gpH are expected to fill the tail tube (Casjens & Hendrix, 1974), a 3 kDA fragment is lost from gpH when the virion is exposed to aggressive protease treatment, suggesting that a section of the protein is exposed to the environment when gpJ is partially cleaved (Roessner & Ihler, 1984).

1.7.4 The phage HK97 as a model system

Phage HK97 is a *Siphoviridae* temperate dsDNA that infects *E. coli* (Dhillon *et al.*, 1980). The T=7 icosahedral head measures ~65 nm in diameter with a flexible tail ~180 nm in length. Analysis of amber mutants allowed for the identification of two assembly intermediates: Prohead I and Prohead II (Duda, Martincic, & Hendrix, 1995; Popa *et al.*, 1991). Between these two intermediates the major capsid protein (mcp) is reduced from 42 kDa to 31 kDa showing that nearly one quarter of the protein is removed proteolytically during assembly (Duda, Martincic, & Hendrix, 1995). In vivo, packaging of the genome triggers expansion followed by crosslinking of the mcp resulting in the mature Head II (H II) structure (Figure 7)(Conway *et al.*, 1995; Dierkes *et al.*, 2009; Duda, 1998; Duda, Martincic, Xie*, et al.*, 1995; Ross *et al.*, 2005). This expansion doubles the internal volume of the capsid (Conway *et al.*, 1995). The identification of these assembly intermediates shows that HK97 follows the same common capsid assembly strategy seen in all other dsDNA tailed phage, including T4 (*Myoviridae*) and P22 (*Podoviridae*), and also in herpesviruses (Casjens & King, 1975)(Casjens and King 1975).

In order to coordinate assembly of the mcps into multimolecular structures effectively, viruses employ scaffolding proteins (Dokland, 1999). However, when expressed alone on a plasmid, the HK97 mcp can self-assemble into the Prohead I structure (Duda, Martincic, & Hendrix, 1995; Oh *et al.*, 2014) while in other systems such as λ, T7 and P22, capsid proteins expressed without their respective scaffolding proteins generate aberrant structures. When HK97 Prohead I capsids are disassembled *in vitro*, they are capable of reassembling in the presence of PEG-8000 and a cation. The reassembled Prohead I particles were visually identical to Prohead I particles assembled *in vivo* suggesting that the mechanism necessary to determine the correct T number is present in the mcp itself. Further, after denaturation, only uncleaved mcp monomers

were capable of refolding and forming capsomers whereas the cleaved mcp was incapable of reassembling into capsomers after denaturation (Xie & Hendrix, 1995). Between the two assembly steps of Prohead I and Prohead II, the N-terminal 102 residues of the mcp are removed–this is termed the 'delta domain' after the analogous region of the T4 major capsid proteins that are absent in the final head structure (Tsugita *et al.*, 1975). When expressed without the delta domain, the capsid protein does not generate Proheads and becomes insoluble, showing that the delta domain is playing the role of a scaffolding protein (Oh *et al.*, 2014). The insolubility of the cleaved capsid protein suggests that the delta domain is required in the initial folding of the capsid protein in conjunction with GroEL and GroES (Ding *et al.*, 1995; Xie & Hendrix, 1995). Further, the delta domain occupies the same position in the genome as most phage scaffolding proteins, between the major capsid protein and the portal gene (Duda, Martincic, & Hendrix, 1995; Duda, Martincic, Xie*, et al.*, 1995; Oh *et al.*, 2014). Altogether, this identified the N-terminal 102 residues as the scaffolding domain for the HK97 capsid assembly system.

Analysis of the HK97 genome identified 3 structural genes required for the assembly of the head: the 47 kDa portal (gp3), the 25 kDa protease (gp4), and the 42 kDa mcp (gp5). The protease is required to remove the delta domain from the mcp to generate the 31 kDa fragment found in the mature capsid (Duda, Martincic, & Hendrix, 1995). The protease has a capsid-targeting signal located in its C-terminus and incorporation into the prohead relies on the N-terminal delta domain α helix (Duda *et al.*, 2013; Krishnamurthy *et al.*, 2014; Oh, 2015). The presence of the protease may play a role in stabilizing the Prohead I structure (Gertsman *et al.*, 2009; Gong *et al.*, 2016; Veesler *et al.*, 2014). The portal is required for DNA packaging and, *in vivo*, the portal may act as the nucleation site for accretion of the hexameric and pentameric capsomers. Mutational studies in the 2nd helical stretch of the delta domain demonstrated that the

delta domain may play a role in the interaction between capsomers and the portal (Oh, 2015). While expansion and cross-linking are triggered *in vivo* by DNA packaging, which requires the portal, expansion and cross-linking can be triggered *in vitro* by lowering the pH and can happen spontaneously at 4°C (Conway *et al.*, 1995; Duda, Hempel*, et al.*, 1995). The relative simplicity of HK97 head assembly makes it an attractive model system. HK97 assembly intermediates and mature particles can be easily isolated by overexpression of the mcp alone or the mcp with the protease. Structures have been obtained for Prohead I, Prohead II, Expansion Intermediates I-IV, and Head II (Hendrix & Johnson, 2012; Lata *et al.*, 2000). Comparing the structures of the proheads to the expanded capsid revealed significant rearrangements of the mcps relative to each other. The prohead is spherical with a bumpy appearance which is the result of the mcps in the capsomers being arranged such that the monomers are overlapping each other significantly resulting in a thick capsid shell. The arrangement of the mcps in the prohead hexamer is 'skewed' such that 3 mcp monomers are shifted relative to the other three. In the expanded capsid, the capsomers have a flatter appearance as the mcps are arranged side-by-side within the capsomers resulting in a thinner capsid shell and symmetric hexamers. When investigating the structures of Prohead I and Prohead II, distinct 'blobs' were observed beneath the capsomers in Prohead I but were not observed in Prohead II (Conway *et al.*, 2007; Conway *et al.*, 1995; Veesler *et al.*, 2014). Since the shell of the capsid can accommodate the atomic model of the mature mcp, which lacks the delta domain, the blobs are likely the delta domains associating in some manner (Gertsman *et al.*, 2009; Huang *et al.*, 2011). When Prohead I particles are heated, the blobs disappear, and the hexamers become more symmetric suggesting that the delta domains are partially responsible for the skewing of the hexamers (Conway *et al.*, 2007).

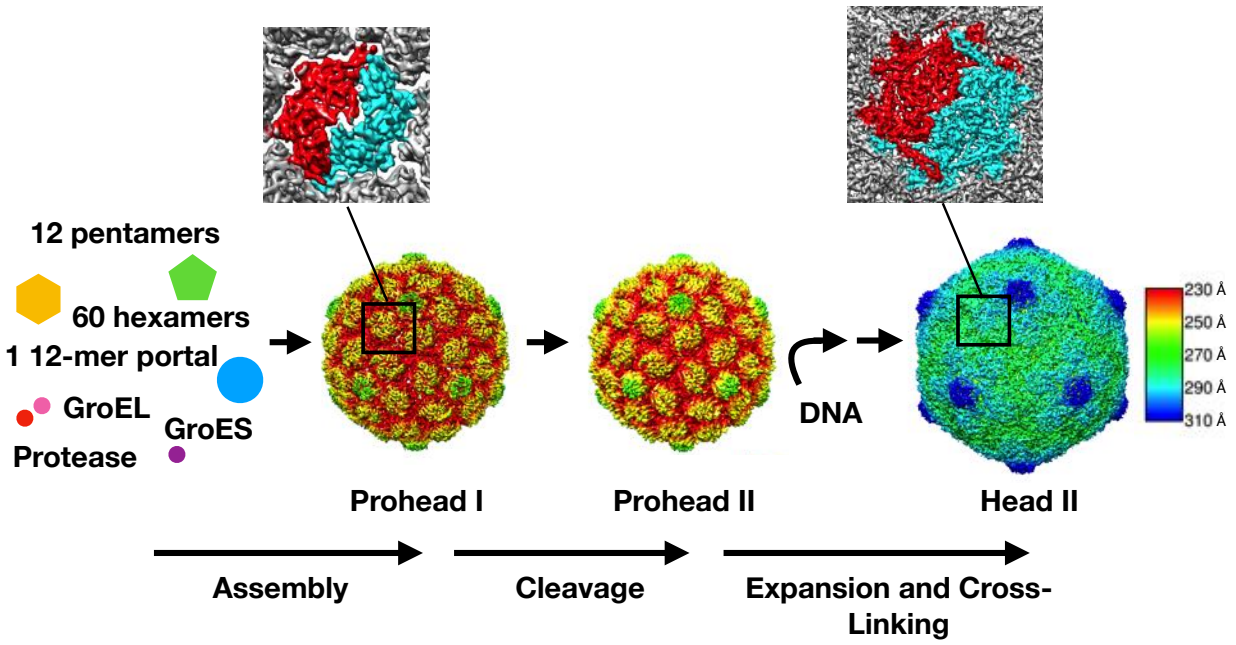

Figure 6. HK97 capsid assembly pathway

Capsid proteins, with the aid of the delta domain, assemble into hexamers and pentamers along with the portal protein and protease. One dodecameric portal complex, 11 pentamers, 60 hexamers, and protease proteins coordinate to form the immature capsid, Prohead I. The protease cleaves the N-terminal delta domain from the capsid protein which then diffuses from the capsid resulting in Prohead II. The dsDNA genome is packaged through the portal triggering expansion and crosslinking resulting in the mature capsid, Head II.

Higher resolution structures revealed the basic elements of the HK97 mcp fold and insights into how the mcps rearrange during expansion (Figure 7). The HK97 mcp has several distinct features: the N-arm; the A (axial) domain that is a triangular structure comprised of α/β elements; the P (peripheral) domain with a substantial "spine helix" and a long B-sheet; and the E (extended) loop which is an anti-parallel β-sheet (Figure 8) (Huang *et al.*, 2011; Wikoff *et al.*, 2003; Wikoff *et al.*, 2000)(Figure 6). Closer analysis of the subunit tertiary structure showed that the alpha helix and beta sheets of the P domain are 'bent' in the Prohead (PDB ID: 3E8K)(Gertsman *et al.*, 2009) and are straightened out in the Head structure (PDB ID: 1OHG)(Helgstrand *et al.*, 2003). Several residue interactions between subunits are constant throughout assembly, thus playing the role of 'staples' to maintain contact and act as pivots throughout the dramatic rearrangement of subunits

and capsomers that occurs during maturation. Analysis of higher resolution structures showed that some interactions like the intracapsomer contact E153/R210 (Hasek *et al.*, 2017) are maintained throughout maturation while others such as intercapsomer contacts K178/D231 (Tso *et al.*, 2014) are only present in the prohead capsid.

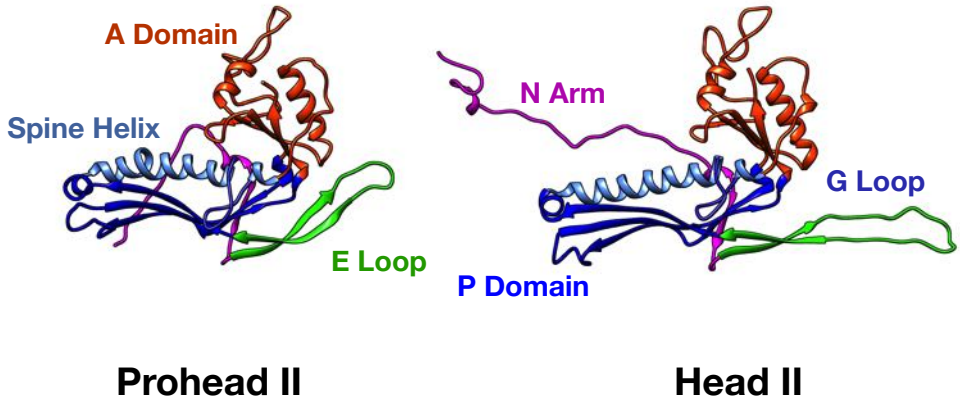

Figure 7. The HK97 mcp fold

The color of the text corresponds to the region of the mcp fold being labeled. PDB IDs: Prohead II 3E8K; Head II 1OHG. (Gertsman *et al.*, 2009; Helgstrand *et al.*, 2003)

Finally, the atomic model of the HK97 mcp revealed structural similarity with other dsDNA tailed phages as well as viruses that infect other domains of life such as archaeal viruses (Pietilä *et al.*, 2013) and Herpesviruses (Baker *et al.*, 2005; Homa *et al.*, 2013; Huet, Makhov*, et al.*, 2016). Further structural similarities were discovered with bacterial encapsulins that form nanocompartments for sequestering enzymes or iron (Akita *et al.*, 2007; McHugh *et al.*, 2014; Sutter *et al.*, 2008). The robustness of the HK97 system and the ubiquity of its fold offer a paradigm for exploring capsid structure and assembly in the vast family of related viruses, including the studies of the phage phi1026b capsid presented in this thesis.

1.8 Cryo-EM use and advances

While the electron microscope has been in use since 1932, cryo-electron microscopy (cryo-EM) wasn't developed until 1982. In the last decade, cryo-EM has undergone a major transformation that is due to advances in camera technology, automation of data acquisition, and reconstruction software (Figure 4). For decades, biological samples were imaged with shadow casting or negative staining, techniques that use a heavy metal or heavy salt to increase the contrast of the sample (Ackermann, 2011; Bradley, 1962). While these techniques aided in visualization, they had the drawback of dehydrating and distorting the samples. Cryo-EM was developed in order to preserve the native environment of the sample by loading the specimen on a holey carbon grid and then plunging the grid into a liquid ethane bath to vitrify the ice (Dubochet *et al.*, 1985; Dubochet *et al.*, 1982).

Over the last 3 decades, advancements in the coherence of the electron beam, camera technology, and image acquisition software has catapulted the technology to new heights (Figure 4) (Chang *et al.*, 2012). Lanthanum hexaboride (LaB_6) crystals improved the brightness and longevity of the electron source over tungsten filament. In the '90s, Field Emission Gun (FEG) technology was applied to transmission electron microscopy using a Tungsten needle (TEM book) resulting in a more spatially coherent point-like source. While charged couple device (CCD) and complementary metal oxide semiconductor (CMOS) cameras were available, film was a primary method of obtaining images from the microscope since the detective quantum efficiency (DQE) at higher frequencies was better than CCD cameras (Bammes *et al.*, 2011; McMullan *et al.*, 2009). In the last decade, the direct electron detector (DED) was developed by several labs focused on increasing the DQE of CMOS based detectors (Guerrini *et al.*, 2011; Mooney *et al.*, 2011). The power of the DED camera was further expanded by the introduction of electron beam-induced

motion correction, a cause of resolution degradation (Li *et al.*, 2013). A major limiting step in reconstructions is the amount of data which depends on image acquisition. One of the first automatic data acquisition software packages, Leginon, was introduced in 2000 for use with CCD digital cameras (Carragher *et al.*, 2000). This software allowed for the collection of hundreds or even thousands of images to be collected during a single experiment. Other software is now also available, including SerialEM (Mastronarde, 2005) and the commercial packages such as FEI's EPU, Gatan's Latitude, and TVIPS' EM-Tools. As the quality and amount of data have both increased, new software analysis tools have taken advantage of high-resolution data resulting in structures with near atomic resolution.

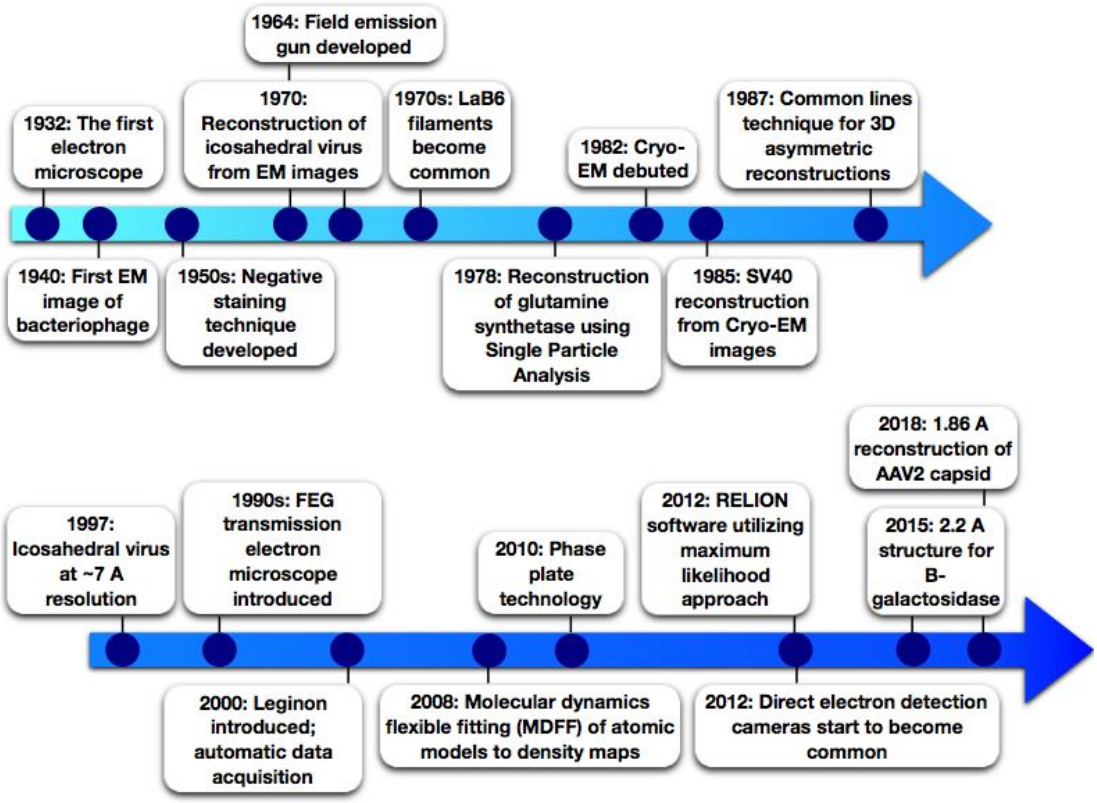

Figure 8. Timeline of achievements and advancements in the field of Cryo-EM

This figure was adapted, altered, and expanded from Frank 2017 (Frank, 2017).

1.9 Thesis Plan

This work utilizes the biophysical method of Cryo-EM to study the structure of capsids and tails for answering questions regarding assembly, structure, and possibly evolutionary relationships. Cryo-EM is particularly well-suited the study of large macromolecular structures such as phage capsids and tails that are comprised of hundreds of copies of individual protein. The overall thesis is organized into two parts: the structure and assembly of phage phi1026b capsids and the structure and assembly of the phage λ tail. Chapter 3 details the high resolution cryo-EM analysis of the phi1026b prohead capsid and head that allows generation of pseudo-atomic models for both capsids. From this work, it is clear that the network of connections critical for assembly of the phage HK97 capsid is altered in phi1026b while features as skewing and rounding of capsomers in the prohead are conserved. Chapter 4 covers the structure and assembly of the λ tail tube. By combining our work with the previously published atomic models of monomeric gpV domains, we are able to detail the structural changes that occur to the λ major tail tube protein during polymerization. Chapter 5 details structural work on the λ tail tip complex (TTC) in both free tails and mature virions. Our model for the TTC of free λ tails suggests that there are structural similarities with the T4 inner baseplate that suggest a strong evolutionary relationship. Finally, Chapter 6 focuses on the presence of proteins throughout λ tail assembly as well as details a putative tail assembly intermediate that is the first such described for a *Siphoviridae* tail.

2.0 Materials and Methods

2.1 Materials

Strains:

E. coli BL21(DE3)-pLysS. Expression strains used for phi1026b proheads and expanded capsids, λ tails, and putative tail assembly intermediates.

E. coli 594. A non-suppressor strain used for complementation tests

E. coli Ymel. A strain used to titer and amplify amber phage and phage.

E. coli C600. A suppressor strain used for complementation tests

Plasmids:

pT7-5-Tail4. A T7 phi10 promoter controls the expression of the tail genes Z, U, V, G, T, H, M, L, K, I, and J (Xu, 2001; Xu *et al.*, 2013, 2014)

pTQ30-G-his-Tail. A digest of pQE30 (Qiagen) that was cloned into the pT7-5, a vector containing a T7 promoter. This plasmid has a N-terminal his-tagged gene G along with the tail genes T, H, M, L, K, I, and J. It was constructed by Jun Xu during his thesis work.

pT7-phi1026b. A T7 phi10 promoter controls the expression of the phi1026b capsid gene, protease gene, and a fragment of the portal gene. This plasmid was constructed by Dr Robert Duda.

Media:

LB Broth: 1% w/v Tryptone, 0.5 % w/v yeast extract and 0.5% NaCl in ddH_2O. When needed, antibiotics were added after the LB broth cooled to reach concentrations of 50 μg/ml Ampicillin and 25 μg/ml Chloramphenicol. In some cases, Maltose was added at 0.4% w/v.

LB Agar: same as LB Broth with 1.5% (w/v) agar added.

MDG non inducing media: 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.2x Trace Metals, 0.5% w/v Glucose, 0.25% w/v Aspartate (Studier, 2005).

Trace metals for MDG media: 20 $\boldsymbol{\mu}$M $CaCl_2$, 10 $\boldsymbol{\mu}$M $FeCl_3$, 10 $\boldsymbol{\mu}$M $MnCl_2$, 10 $\boldsymbol{\mu}$M $ZnSO_4$, 0.1 $\boldsymbol{\mu}$M $CoCl_2$, 0.05 $\boldsymbol{\mu}$M $CuCl_2$, 0.05 $\boldsymbol{\mu}$M $NiCl_2$ (Studier, 2005).

Buffers:

Lambda Dilution Buffer: 10 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$

TKG: 20 mM Tris-HCl, pH 7.5, 100 mM potassium glutamate

Lysis Buffer: 50 mM Tris-HCl, pH 8.0, 5 mM EDTA

Buffer DL: 20 mM Tris HCl, 40 mM NaCl (pH 7.5)

Buffer G: 20 mM Tris HCl, pH 7.5, 100 mM NaCl

TAMg Buffer: 40 mM Tris base, 20 mM acetic acid pH 8.1, 1 mM magnesium sulfate

Gel Buffers:

50X TAE: 1 M Tris-acetate, 50 mM EDTA pH 8.0

4X Lower Buffer for SDS-PAGE: 1.5 M Tris-HCl pH 8.8, 0.4% SDS

4X Upper Buffer for SDS-PAGE: 0.5 M Tris-HCl pH 6.8, 0.4% SDS

10X Running Buffer for SDS-PAGE: 0.25 Tris base, 2.5 M glycine, 1% SDS

4X Sample buffer for SDS-PAGE: 0.25 M Tris-HCl pH 6.8, 40% glycerol, 20% B-mercaptoethanol, 8% SDS

Nickel column buffers (NOVAGEN):

1X Charging Buffer: 50 mM $NiSO_4$

1X Binding Buffer: 20 mM Imidazole, 0.5 M NaCl, 20 mM Tris-HCl pH 7.9

1X Wash Buffer: 80 mM Imidazole, 0.5 M NaCl, 20 mM Tris-HCl pH 7.9

1X Elute Buffer: 1 M Imidazole, 0.5 mM NaCl, 20 mM Tris-HCl pH 7.9

1X Strip Buffer: 100 mM EDTA, 0.5 M NaCl, 20 mM Tris-HCl pH 7.9

2.2 Protein purification and analysis

Native Agarose Gel Electrophoresis

Native agarose gels were made by diluting agar to a final concentration of 1% w/v in 1x TAMg buffer. Protein samples were mixed with a TAMg buffer dye/glycerol mix (40 mM Tris base, 20 mM acetic acid pH 8.1, 1 mM magnesium sulfate). Gels were run at a contant voltage of 125 V till the sample dye had move towards the end of the gel. Gels were stained with Coomassie Brilliant Blue R-250 (0.05% w/v Coomassie Brilliant Blue R-250, 50% v/v methanol, 10% v/v acetic acid)

SDS gel electrophoresis

Samples were mixed with the appropriate amount of 4x SDS sample buffer and boiled for 2 minutes (Laemmli, 1970). Samples were then loaded on a 12% low-crosslinking SDS polyacrylamide gel (33.5% w/v acrylamide/0.3% w/v methylene bisacrylamide) (Laemmli, 1970). The gel was run at a constant voltage of 125 V till the sample dye reached the end of the gel (approximately 45 minutes). Gels were stained with Coomassie Brilliant Blue R-250 (0.05% w/v Coomassie Brilliant Blue R-250, 50% v/v methanol, 10% v/v acetic acid)

Prohead Purification

Cells were grown overnight in MDG then diluted into 300 ml TYM5052 with 50 μg/mL ampicillin and 25 μg/mL chloramphenicol. Cells were then shaken at 300 RPM at 37° C for 24 hours and then pelleted in a JA16.250 rotor at 6,000 RPM for 10 minutes. The pellet was resuspended in 60 ml lysis buffer. Lysis was facilitated by the addition of Triton X100, DNaseI, $MgSO_4$, and temperature cycling. Cell debris were removed by pelleting in a JA16.250 rotor at 8000 RPM for

12 minutes. Proheads were isolated from the supernatant with polyetheyleneglycol (PEG) precipitation by adding NaCl to a final concentration of 0.5 M and PEG 8000 to 7% (w/v). PEG precipitate was isolated by centrifugation in a JA16.250 rotor at 8,000 RPM for 12 minutes. The pellet was further cleaned by resuspending the pellet in 30 ml Buffer G and centrifugation. The supernatant was then diluted with Buffer G and proheads were pelleted using ultracentrifugation in a Ti45 rotor for 1 hour and 45 minutes at 35,000 RPM. The pellets were resuspended in Buffer G overnight. Any insoluble material in the pellet was moved by centrifugation in a JA18 rotor at 8000 RPM for 10 minutes. The supernatant was then loaded on a 10-30% v/v glycerol gradients which were then spun in a SW28 rotor for 1 hour and 30 minutes at 27,000 RPM in the cold. Prohead bands were removed with a 18 gauge needle. Bands were dialyzed against 5 mM Tris pH 7.5 to remove glycerol. Ion exchange chromatography was performed using a HQ20 column for anion exchange on the BioCad Workstation (PerSeptive Biosystems, Framingham MA?). Fractions containing peaks were concentrated by centrifugation in a Ti80 rotor at 50,000 RPM for 2.5 hours. Pellets were resuspended in a small volume overnight.

λ tail protein expression

Plasmids with the T7 promoter were expressed in BL21(DE3)plysS cells (Studier, 2005; William Studier *et al.*, 1990). A colony was placed in 1.5 LB Amp/Cam (50 μg/ml Amp, 25 μg/ml Cam) in a tube and left to shake overnight at 37°. The overnight culture was diluted 1:2000 into 300 ml or 1000 ml. The cells were grown at 37° at 300 RPM shaking till the OD_{550} reached 0.3 and induced with IPTG at a final concentration of 0.4 mM. Expression was allowed to continue for 3 hours.

Tail Purification

The desired genes on plasmids listed above were expressed in BL21(DE3)pLysS cells using the method described above. After expression, samples were spun in a JA 10 rotor at 6,000 RPM for 10 minutes, resuspended in 60 ml lysis buffer and lysed. Cell debris was removed from the lysate by centrifugation in a JA25.50 rotor at 13,000 RPM for 14 minutes Tails were isolated by the addition of potassium glutamate to 0.7 M and 7.5% w/v PEG 8000 to facilitate precipitation. After 30 minutes, the PEG precipitate was isolated by centrifugation in a JA25.50 rotor at 11,000 RPM for 12 minutes. 2 to 3 ml of the resuspended pellet was loaded onto a 10 to 30% glycerol gradient and centrifuged in a SW28 or SW32 rotor at 27,000 RPM for 5 hours at 4° Celsius. Bands were identified using a light source and removed with a 20G needle. Tail fractions were then diluted in TKG buffer for ultracentrifugation in a Ti45 rotor at 40,000 RPM for 4.5 hours. The resulting pellet was then resuspended in a small volume overnight. Ion exchange chromatography was performed using a HQ20 column for anion exchange on the BioCad Workstation (PerSeptive Biosystems, Framingham MA). Fractions containing peaks were concentrated by centrifugation in a Ti80 rotor at 50,000 RPM for 2.5 hours. Pellets were resuspended in a small volume overnight.

Nickel Affinity Chromatography

Nickel affinity chromatography was performed using Novagen protocols (Xu, 2001). The column was rinsed of strip buffer by several column volumes of diluted water. The column was charged with 4 column volumes of charging buffer then equilibrated with 3 column volumes of binding buffer. Once the sample was loaded onto the column, the column was washed with 10 column volumes of binding buffer. Eight column volumes of washing buffer were then passed through the

column followed by 3 bed volumes of elute buffer. The column was then cleaned with 5 column volumes of strip buffer and then stored.

Negative Stain Transmission Electron Microscopy

Continuous carbon grids were glow charged for a minimum of 45 seconds using an EmiTech K100X Glow Discharger. Three µL of sample was pipetted onto freshly glow-discharged grids and left standing for 15 seconds before blotting excess sample without drying the sample. Grids were then washed with water, blotted, stained with 1% w/v uranyl acetate, blotted again, and left to air-dry. Grids were imaged using either an FEI Tecnai T12 microscope (FEI, Hillsboro, OR) with a 1024x1024 pixel Gatan US1000 CCD camera (Gatan, Pleasanton, CA) or an FEI Tecnai TF20 microscope with a 4096x4096 pixel Gatan US4000 CCD camera.

Cryo-Electron Microscopy

Depending on the experiment, the grids used were Quantifoil R2/1 copper grids (Quantifoil Micro Tools GmbH, Jena, Germany) or C-Flat CF-2/1-4C (Protochips Inc, Morrisville, NC), both with 2µm holes spaced 1µm apart in a carbon film. Grids were glow discharged for 10 seconds using an EmiTech K100X Glow Discharger (Quorum Technologies, Ashford, Kent, UK). 3.5µL of sample was deposited on Quantifoil grids, but ≤ 2.5 µL of sample were deposited on C-Flat grids. Grids were cryo-plunged into liquid ethane or a mixture of liquid ethane and propane (Tivol *et al.*, 2008) using an FEI Vitrobot Mark III cryo-plunger. During the course of the work, grids were initially single-blotted before plunging, but ice thickness was found to be more even by double-blotting grids with the Vitrobot chamber set to 100% humidity and 4°C. Double-blotting lasted a

total of 7 to 8 seconds with an automatic rotation of the blotting paper between blots so that each blot was with a fresh area of filter paper.

Cryo-negative stain TEM

Cryo-negative staining was performed using a slurry of ammonium molybdate (Adrian *et al.*, 1998; De Carlo *et al.*, 2008; De Carlo *et al.*, 2002; De Carlo & Harris, 2011). The ammonium molybdate stain was made by diluting 1.25 g ammonium molybdate in 875 µL ddH_2O and 125 µL 10 M NaOH. Prior to use, the slurry was thoroughly vortexed. Once the slurry started to separate, ~4 µL of stain was removed from the top of the tube and mixed with ~1 µL sample. Approximately 4 µL of mixture is placed on a freshly glow discharged grid. Grids were blotted with an FEI Vitrobot Mark III for 7.5 seconds at a ~70% humidity at room temperature. After blotting, the grids were allowed to 'drain' for 5 to 6 seconds (ie, a delay) prior to plunging. Double blotting, as described above, was not used when generating cryo-negative stain grids.

Data collection

Frozen grids were loaded into a FEI Krios microscope (phage phi1026b capsids) or a FEI Polara microscope (phage λ tails and virions). Both microscopes were operated at 300 kV and equipped with FEI Falcon 2 direct electron-detecting (DED) cameras, and data collection was performed using the FEI EPU automation software. The phage phi1026b capsid sample was imaged at a nominal magnification of 75,000x with a post-column magnification of 1.6x, corresponding to a calibrated pixel size of 1.08 Å at the sample. Movie mode was not an option at the time this dataset was collected, so the micrographs were not drift-corrected. The phage λ samples were images at a nominal magnification of 93,000x with a post-column magnification of 1.4x, corresponding to a

calibrated pixel size of 1.15 Å at the sample. Movie mode was used in the phage λ data collection with a 7-frame limit imposed by the Windows XP host PC that was nonetheless effective for correcting modest drift.

2.3 Structure determination, modeling, and visualization

Capsid structure

Estimates of defocus and astigmatism for each micrograph were done manually with the BSOFT suite (Heymann & Belnap, 2007). Over 70,000 particle images of phage phi1026b capsids were selected and extracted from micrographs using the x3d program (Conway & Steven, 1999). A starting model was generated ab initio with the Random Model (RM) method (Yan *et al.*, 2007). Structures were further refined using the software package AUTO3DEM (Yan *et al.*, 2007). Reconstructions were continued once the "flattenmap" feature was introduced in an AUTO3DEM update. Class averaging was performed in the software suite EMAN2 (Bell *et al.*, 2016; Tang *et al.*, 2007) to investigate whether the particle population was homogenous or not. An ad hoc program, ClassifyReproj (Conway, unpublished), was developed to classify particles based on differences in density internal to the capsid that may be specific to Prohead I capsid forms (including the sizeable N-terminal delta-domain) compared to the Prohead II capsids where this domain has been proteolyzed and expelled (Conway *et al.*, 1995).

Tail tube Structure

Tail tube images were selected and extracted from micrographs using the e2helixboxer.py program in the EMAN2 software suite. Tail tube structures were determined using three different software approaches for comparison.

Spring:

Micrographs and tail tube coordinates obtained from e2helixboxer.py were imported into the Spring software suite (Desfosses *et al.*, 2014). Estimates of defocus were performed in the spring software suite using MICCTFDETERMINE which uses CTFFIND (Mindell & Grigorieff, 2003). Particles were generated from image coordinates using SEGMENT. SEGMENTEXAM, SEGCLASSRECONSTRUCT, and SEGGRIDEXPLORE were used to identify degree of out-of-plane tilt and helical parameters of the particles. Starting models were generated within the SEGMENTREFINE3D program. Structures were B-factor sharpened and filtered to a resolution cutoff using SEGREFINE3DINSPECT.

Relion:

CTFFIND defocus values obtained from CTFFIND3 and particle coordinates obtained using the EMAN2 e2helixboxer.py program, and then imported directly into Relion 2.0 (Scheres, 2012), which includes helix processing. 3d refinements were performed both with a starting model (from spring, as above) and without. Refinements were run with and without fixed helical parameters. When helical parameters were fixed, they were fixed at a rise of 42.8 A and a rotation of 18.13°. B factor sharpening was performed using relion_postprocess.

IHRSR++:

An image stack generated in EMAN2 using e2helixboxer was converted into the Spider format (Frank *et al.*, 1996). Helical parameters were obtained from averaged power spectra by Alexis Huet. Reconstructions were performed on particles with a box size of 400 x 400 pixels.

Tail tip complex

Reconstructions of the λ tail tip complex were performed using several software packages and pathways. Estimates of defocus and astigmatism were performed automatically using CTFFIND3 or the ctffind plug-in within the EMAN2 software suit. Particle images were selected from micrographs using the x3d program (Conway and Steven 1999). A starting model was generated ab initio with the Random Model (RM) method in AUTO3DEM and applying either no symmetry, 3-fold symmetry or 6-fold symmetry about the tail axis. When performing reconstructions in AUTO3DEM, particle orientations were set to 90° from the azimuthal plane to aid in alignment by manually altering the particle coordinate files (dat files). Multiple refinements were done with varying symmetries and box sizes. Box sizes ranged from 150 x 150 pixels to 400 x 400 pixels. Particle coordinates were imported into EMAN2 and Relion to continue refinement of the reconstructions using the AUTO3DEM model as a starting point. 2D class averages were performed separately in EMAN2.

Atomic modeling and visualization

Phenix:

Density maps of the phi1026b capsid were converted to the XPLOR format using the EMAN2 e2proc3dy program, then converting to the "map" format using the mapman and mapmask programs (Kleywegt & Jones, 1996) included in the CCP4 software suite (Winn *et al.*, 2011). The resulting map was then converted to structure factors using phenix.map_to_structure_factors in the Phenix software suite (Adams *et al.*, 2010). Developments in Phenix during the analysis included the real_space_refine function, for which the cryo-EM density maps were converted into the ccp4 format using the EMAN2 program e2proc3d.py. Alignment of the atomic model and the map were confirmed in the COOT software package (Emsley *et al.*, 2010) – see below. Phenix.real_space_refine was run with minimization, morphing, and simulated annealing.

Atomic model manipulation in COOT:

Additional modelling of the phi1026b N-arm was performed in COOT. Residues were built on to the model using features within COOT. Localized fitting was performed with the 'real space refine zone' feature and irregularities in geometry were fixed using the 'regularize zone' feature.

MDFF/IMD-MDFF:

Flexible fitting of atomic models into cryo-EM density maps was performed using the molecular dynamics flexible fitting (MDFF) approach (Trabuco *et al.*, 2008; Trabuco *et al.*, 2009). Two protocols were followed depending on the stage of fitting and needs. The first was performed without a solvent by following the MDFF protocol for flexible fitting performed in vacuum and

intended for large scale manipulation of the structure. Interactive molecular dynamics (IMD-MDFF)(Stone *et al.*, 2001)as used when either a large alteration in the atomic model was needed or one section of the fold needed to be altered. When focusing on one section of a fold, the rest of the atomic model was fixed to prevent changes to the model as well as to reduce computational load. This approach was used for initial fitting of the phi1026b mcp atomic model and the fitting of the gpV N-terminmal tube domain atomic model to their respective densities. The second protocol was a molecular dynamics simulation performed with a solvent so that the structure could adjust to the cryo-EM density both without restraints and in a more natural environment. This second protocal was used with the phi1026b Prohead I mcp atomic model. To generate an atomic structure in an ionized waterbox, the MDFF tutorial was followed. The protocol involves two minimization steps and two simulation steps for a total of 4 steps. The first constrained minimization for 5000 minsteps at a gscale of 0.3 at a temperature of 300 K with the peptide constrained to allow relaxation of the solvent. The peptide was constrained with a constant of 450 kcal/A^2. The second minimization was a full minimization of both the solvent and the peptide for the same values of minsteps, gscale and temperature. The first simulation was a partially constrained simulation that was started at a temperature of 0 K and ramped to 310 K. The solute was constrained with a constant of 10 kcal/A^2. The final simulation was unconstrained and run at the temperature 300 K.

Chimera/VMD:

Maps for use in Phenix and MDFF were generated in UCSF Chimera (Eric F Pettersen *et al.*, 2004). Atomic models were fit into density maps as rigid bodies with the ‘fit in map’ feature. Density maps were also compared after rigid fitting to access structural differences between

reconstructions. Density selected in a zone around a fitted atomic model was further edited manually to erase contributions from an adjoining monomer. Superimposing structures was achieved using the Matchmaker plugin (Meng *et al.*, 2006). Planes defined by selecting three corresponding atoms in adjacent capsomers were used to estimate angles between the capsomers. Electrostatic surface potentials were generated using the PDB2PQR and APBS plugins. Visualization of MDFF trajectories were performed in VMD (Humphrey *et al.*, 1996) and then further observed in Chimera.

SPDBV:

Threading of the phi1026b major capsid protein's amino acid sequence onto the atomic models for the HK97 Prohead I structure and expanded capsid structure was performed using the Swiss-Model server and Swiss PDB Viewer (Guex & Peitsch, 1997).

3.0 The Capsid Structure of Phage phi1026b

3.1 Introduction

In capsid assembly, individual proteins coordinate to generate capsid structures with consistent shape and size (Figure 9A). While in tail assembly, bacteriophages utilize a "tape measure" protein to regulate tail size (length) (Katsura & Hendrix, 1984), no mechanism for determining capsid size has been identified. The majority of phages that have been studied generate unique structures with the same uniformity as tail assembly. The common assembly pathway involves capsomers, pentameric and hexameric complexes of the mcp, pentons and hexons, forming the prohead with the aid of scaffolding proteins along with the dodecameric portal occupying one vertex of the capsid. The prohead is the immature form of the capsid. The scaffolding proteins either diffuse out of the capsid or are removed by a phage encoded protease to form the empty Prohead II capsid. DNA packaging occurs through the portal triggering expansion, which is an overall increase in the diameter of the capsid facilitated by the rearrangement of the capsomers (Conway *et al.*, 1995). Then, any additional structural proteins or modifications which stabilize the capsid can occur. At this point, the tail attaches which to form the mature virion.

The phage HK97 major capsid protein (mcp) assembles into Prohead I structures with the icosahedral triangulation number T=7 in the absence of additional proteins, suggesting that size determination resides completely within the mcp itself. As discussed earlier, the N-terminal delta domain acts as a scaffolding protein but what role it plays in determining capsid size is unclear. Removal of residues 4 through 13 at the N-terminus, yielded regular T=7 Prohead I capsids (Oh

et al., 2014), but deletion of 7 residues from the middle of the delta domain produced a mix of T=4 Prohead-like capsids as well as T=7 Proheads (Oh, 2015). Further, the delta domain of the HK97-like phage T5 mcp is 57 residues longer than that of HK97 and the T5 capsid is larger, with a T=13 icosahedral lattice. In phage T4, mutations in the separate scaffolding core protein, gp22, result in capsids with altered width, symmetry, or multiple tails (Keller *et al.*, 1988). Together, this evidence suggests that size is determined by scaffolding domains. However, mutations in the mature domain of the mcp or in other capsid proteins can also result in changes to capsid size and shape. In phage T4, a single point mutation in the mcp gp23, that shares structural similarity with HK97, yields isometric capsids in addition to wild type prolate capsids (Chen *et al.*, 2017). Altogether, these data suggest that the delta domain or the scaffolding protein works with the mature mcp domain to determine size and shape. Given the vast family of dsDNA tailed phages employing the HK97 fold, studying a virus highly similar to HK97 that assembles a capsid of a different size may help illuminate how these two functional domains work together to determine capsid size.

phi1026b is a temperate bacteriophage of *Burkholderia pseudomallei* strain 1026b that infects *Burkholderia mallei,* a pathogen that causes meliodosis, also called Whitmore's disease, that is endemic to Thailand, Vietnam, and Northern Australia (DeShazer, 2004). The capsid of phi1026b is ~56 nm in diameter and has a non-contractile tail that is ~200 nm long and 8 nm in diameter. phi1026b has the same number of capsid genes as HK97 with no apparent separate scaffolding protein. Also, the mature domain of the mcp shares 50% sequence similarity to that of the HK97 mature capsid domain, including conserved crosslinking residues and the delta domain cleavage site (Figure 9C). One difference is that the delta domain is 22 residues longer than that of HK97. Further, the phi1026b genome is 54,865 bp, which is 38% larger than that of HK97 (39,732 bp), strongly suggesting that the capsid is larger. Since the phage phi1026b mcp is very

similar to the HK97 mcp but may have a larger capsid, we chose to study phi1026b capsid assembly to see how closely related mcps assemble capsids of different sizes.

This chapter aims to compare the capsids of HK97 and phi1026b. Comparing the two structures may give us insights into how the HK97 fold is adapted to generate phage of different sizes.

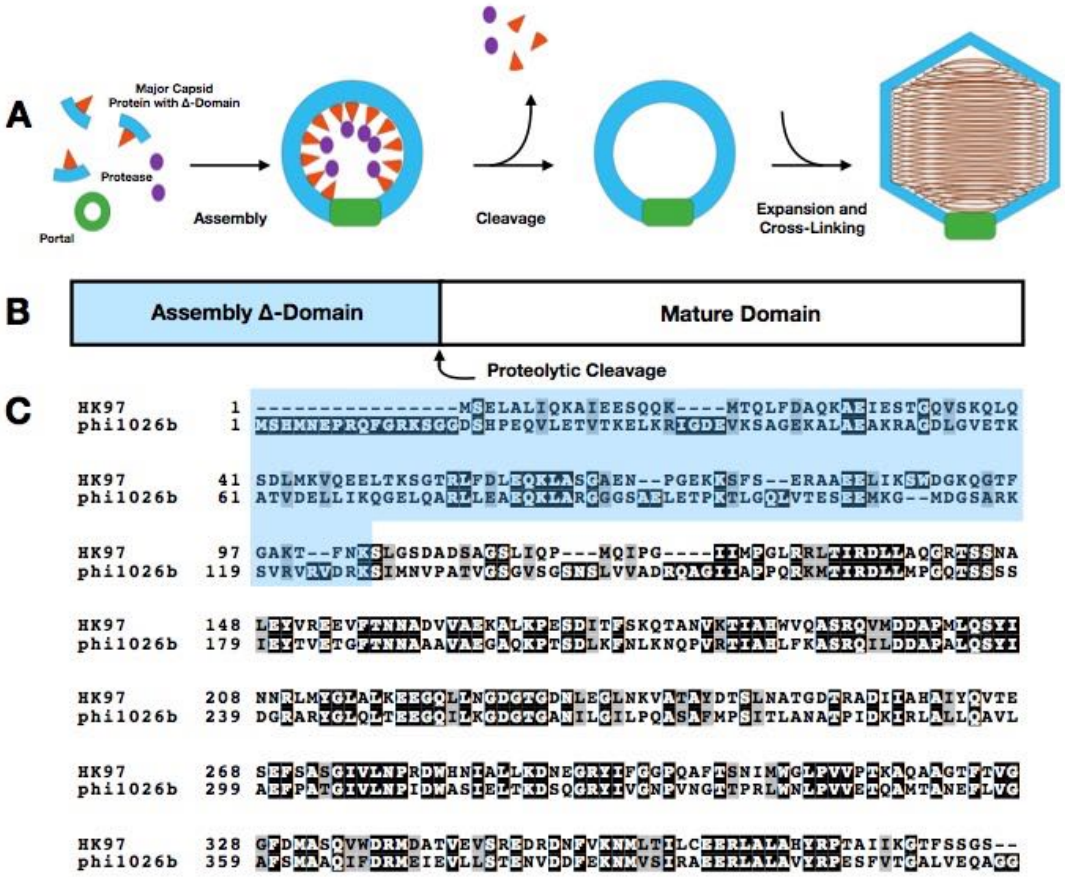

Figure 9. Representatives of capsid assembly and capsid protein alignment

A) Model of the HK97 capsid assembly pathway. Capsid proteins with the catalytically attached delta domain along with the portal protein and protease assemble into hexamers and pentamers. 1 dodecameric portal complex, 11 pentons, 60 hexons, and protease proteins coordinate to form Prohead I. The protease cleaves the N-terminal delta domain from the capsid protein resulting in Prohead II. The dsDNA genome is packaged leading to expansion and crosslinking resulting in the mature Head II structure. B) Model of the two domains of HK97 capsid proteins showing the location of the proteolutic cleavage site. C) The alignment of the capsid proteins from HK97 and phi1026b with the location of the delta domain highlighted in blue.

3.2 Results

3.2.1 Expression of phi1026b mcp and protease genes yields particles

Dr. Robert Duda purified the proheads and expanded capsids of phi1026b for this work. The cloned genes for the phi1026b capsid proteins were provided by Dr. David DeShazer (DeShazer, 2004). The major capsid protein and protease genes were expressed from a plasmid containing a T7 promoter in the *E. coli* strain BL21(DE3)pLysS (Studier, 2005; William Studier *et al.*, 1990). A detailed protocol for the purification of particles is outlined in the Materials and Methods section. Four bands were observed in the glycerol gradients (Figure 10A) used as the major step in the purification. Samples taken after successive steps in the purification analyzed on native agarose gels are shown in Figure 10B,C. The samples had bands that correspond to proheads, expanded capsids and capsomers. The major band found in the glycerol gradients contained the fastest sedimenting particles and the particles in this band were further purified using ion exchange chromatography (IEX). The IEX fractions that contained the major peak fractions were pooled, concentrated and used for TEM imaging, which showed that this preparation contained proheads with visible structures in their interiors and expanded capsids (Figure 10E). SDS gels showed a band that corresponds to the uncleaved mcp protein as well as a second band that corresponds to the mcp with the N-terminal 13 residues removed (Duda unpublished N-terminal sequencing results) in the final preparation (Figure 10D), suggesting that the proheads observed in the TEM images are Prohead I-like. A band corresponding to the protease was also present in most samples. Unexpectedly, this protease band was clearly present in the final preparation of proheads. The protease was expected to cleave all of the proheads to produce fully cleaved Prohead II, but its activity was apparently very weak or compromised (perhaps due to

expression in a heterologous host). We conclude that the final preparation contained mostly uncleaved Prohead I containing a non-active protease and a few expanded capsids (Figure 10F).

A minor band at ~31 kDa, which is the expected size of the cleaved mcp, was present in the first, second, and third glycerol gradient bands but barely visible in the fourth band that we conclude contained mostly Prohead I. The presence of the 31 kDa cleaved mcp indicates that some genuine Prohead II was made, but apparently all of it rapidly converted to fully expanded and crosslinked heads during subsequent steps of purification. We expected to observe high molecular weight bands corresponding to crosslinked phi1026b mcp in the samples that contain expanded heads, but apparently the crosslinking is so efficient in phi1026b capsids that nearly all of the protein was converted into forms so large that they do not enter the SDS gels, as expected for the large topologically intertwined protein networks (Duda, 1998) that form when phage HK97 mcps undergo crosslinking.

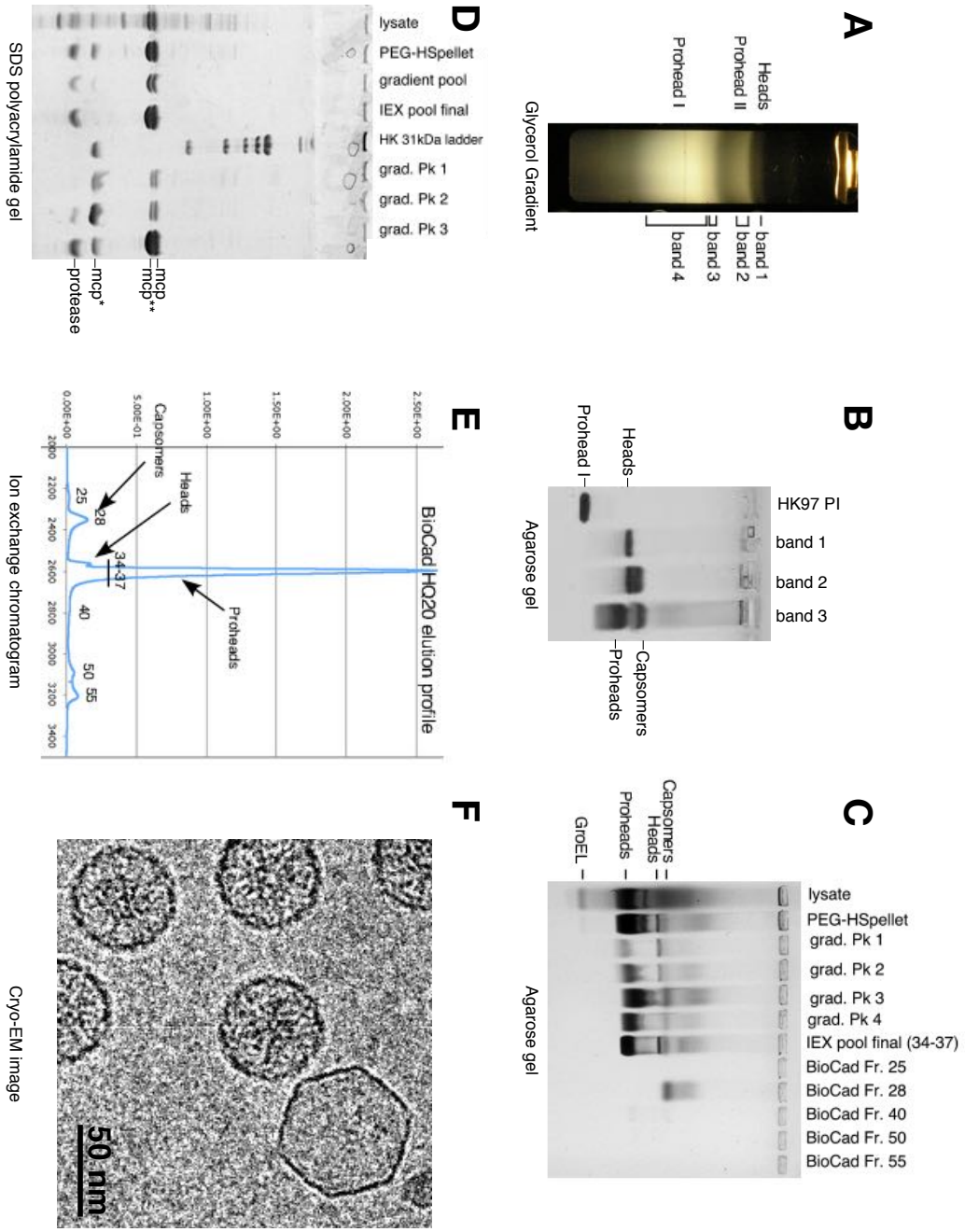

Figure 10. Purification of phi1026b capsids

A) An image of a glycerol gradient is shown containing 4 bands which are marked to the right of the image. The types of capsids inferred to be the major component of each bands is marked to the left of the image. B) An agarose gel containing pelleted samples from the glycerol gradient is shown. A HK97 Prohead I sample is included for reference. C) An agarose gel containing samples from several purification steps is shown. D) A SDS polyacrylamide gel containing samples from purification steps is shown. E) A chromatogram from the IEX purification of the particles present in the major gradient band (band 4) shown. The inferred contents of several peak are marked on the image. F) A cryo-EM image of the IEX final fraction showing both proheads and expanded capsids is shown.

3.2.2 The T=9 structures of the phi1026b Prohead I and Head

I collected images of over 70,000 particles from a data set obtained on a FEI Krios factory microscope in Eindhoven, The Netherlands, with the help of Dr. Matthijn Vos. I endeavored to gather as many particles as possible including those present on micrographs with considerable drift and particles cut off at the edge of the image. A map of ~5 Å resolution was obtained for the phi1026b Prohead structure (Figure 11B, top, D). Later, a ~6 Å map was generated for the expanded capsid structure by Dr James Conway (Figure 11C, bottom, D). Both the prohead and expanded capsid structures show icosahedral (T=9) caspids with α-helices and β sheets resolved as well as many bulky side chains. Resmap calculations (Kucukelbir *et al.*, 2014) show that a majority of the prohead capsid shell is at a resolution of ~5.2 Å (Figure 13A) and flexible regions of the mcp such as the N-arm, A-domain and E-loops are at a lower resolution of ~8 Å (Figure 13B).

The phi1026b prohead and head capsid structures share the same overall morphology that is found in the HK97 prohead and head structures (Figure 11A, B). The phi1026b Prohead structure has the rounded structure typical of early capsid assembly intermediates and from vertex-to-vertex, the capsids have a diameter of 61 nm. The expanded capsid has the angular appearance typical of fully expanded capsids. As measured in the cryo-EM reconstruction, the vertex-to-vertex distance is ~76 nm in diameter consistent with the cryo-EM micrographs. Similar to the HK97 Prohead and expanded capsid structures, the phi1026b Prohead has thicker capsid walls than the phi1026b expanded capsid. The phi1026b Prohead I capsid has walls ~60 A thick whereas the expanded capsid has walls ~25 A thick. The micrographs showed that the Proheads have internal structures whereas the expanded capsids do not and this bears out in the reconstructions.

The phi1026b capsids are composed of 12 pentons (as the portal is not present) and 72 hexons arranged in T=9 icosahedral symmetry. In Prohead I, there are 12 hexons on the 3-fold axis (Figure 14A-magenta) and 60 hexons that surround the pentons (Figure 14A-purple). The hexons sitting on the 3-fold axis appear symmetric since they feature 3-fold symmetry whereas the other 60 hexons appear asymmetric. The terms 'symmetric' and 'asymmetric' are used to differentiate between the two. The asymmetric hexons display 'skewing' that is similar to the skewing observed in HK97 prohead hexons (Figure 11A, B-insets). In both phages, the skewing is a result of a shift of three monomers of mcp within the hexon relative to the other three monomers resulting in pseudo 2-fold symmetry (Figure 14B). The two groups of three monomers are also rotated out of plane relative to each other such that one side of the skew is lifted higher than the other side. The symmetric hexon also features a form of skewing where instead of two groups of monomers are shifted relative to each other, it is three groups of monomers resulting in a 'twisted skew' (Figure 14C). This results in what appears as three wedges that are oriented so that one side of the wedge is 'below' an adjoining wedge on one side and 'above' an adjoining wedge on the other side. Collectively the monomers of the prohead structure are arranged such that the capsid surface has a 'blistered' appearance which is also seen in HK97. In the expanded capsid, the two hexons are nearly identical in size and the subunits of the capsomers are arranged such that the capsid has a flat surface. A similar change in orientation of the capsid monomers between the two states is also observed in the HK97 capsid assembly pathway.

The internal density in the phi1026b Prohead structure is similar to the HK97 Prohead I structure. In the micrographs, the phi1026b Proheads appear to have internal density and these are resolved as columnar density beneath the capsomers in the reconstruction. These columns appear at the approximate center of the capsomers and protrude towards the center of the capsid (Figure

11B, top). The HK97 Prohead I density map also has columnar density beneath the capsomers but they are smaller than those of phi1026b (Figure 11A, top). The same columnar density is also observed in the Prohead I structure for phage T5 (Huet, Duda*, et al.*, 2016).

Fitting of HK97 PDB models into the phi1026b density maps confirms that the phi1026b mcp features the HK97 fold. In Figure 12A, the HK97 Prohead I crystal structure assymetric unit (PDB ID: 3QPR) has been rigid body fitted into the phi1026b Prohead I EM density map. In Figure 12C, the HK97 Head crystal structure assymetric unit (PDB ID: 1OHG) has been rigid body fitted into the phi1026b head density. The A-domains, spine helices, P-domains and G-loops in both the Prohead I and expanded capsid are easily recognizable in the maps. In Prohead I map, the ends of the E-loops are not visible (very likely due to their flexibility as is the case for HK97 proheads). In the expanded capsid, the E-loops are visible and in the down position that is required for cross-linking in HK97 (Gan *et al.*, 2006). The A domains of the HK97 Prohead I crystal structure do not align within the A domains of the phi1026b map (Figure 12A, inset), showing that the relative orientation of the subunits within the hexon are different. The A-domains of the HK97 Head crystal structure fit neatly into the phi1026b density but the P domains and E-loops of the HK97 crystal structure do not (Figure 12C, inset). It appears that the P domains and E-loops of the phi1026b expanded capsid capsomers are oriented further away from the center relative to that of the HK97 expanded capsid capsomers. When individual monomers HK97 Prohead I mcp and HK97 Head mcp are rigid body fitted into the density the misalignment is not as obvious (Figure 12B, D).

In both instances, flexible fitting was required to fit HK97 fold to the phi1026b cryo-EM map. The sequence for the mcp protein was threaded onto the PDB models for the HK97 Prohead-1 and expanded capsid using SPDBV and the Swiss Model Server (Guex & Peitsch, 1997). The Prohead I model was initially fit into the density using molecular dynamics flexible fitting (MDFF)

in vacuum (Trabuco *et al.*, 2008; Trabuco *et al.*, 2009). Further fitting was performed in a waterbox with 100 mM NaCl to better approximate in vitro conditions. This was performed by building up the asymmetric unit and surrounding monomers through successive molecular dynamics trajectories (personal communication Dr. Zuckerman, Dr. Suarez, Dr. Pedilla, Dr. Schulten). A more detailed description of this protocol can be found in the Materials and Methods. Areas of the fold that did not occupy density after fitting were manually placed into the density using interactive molecular dynamics (IMD-MDFF) in vacuum (Stone *et al.*, 2001).

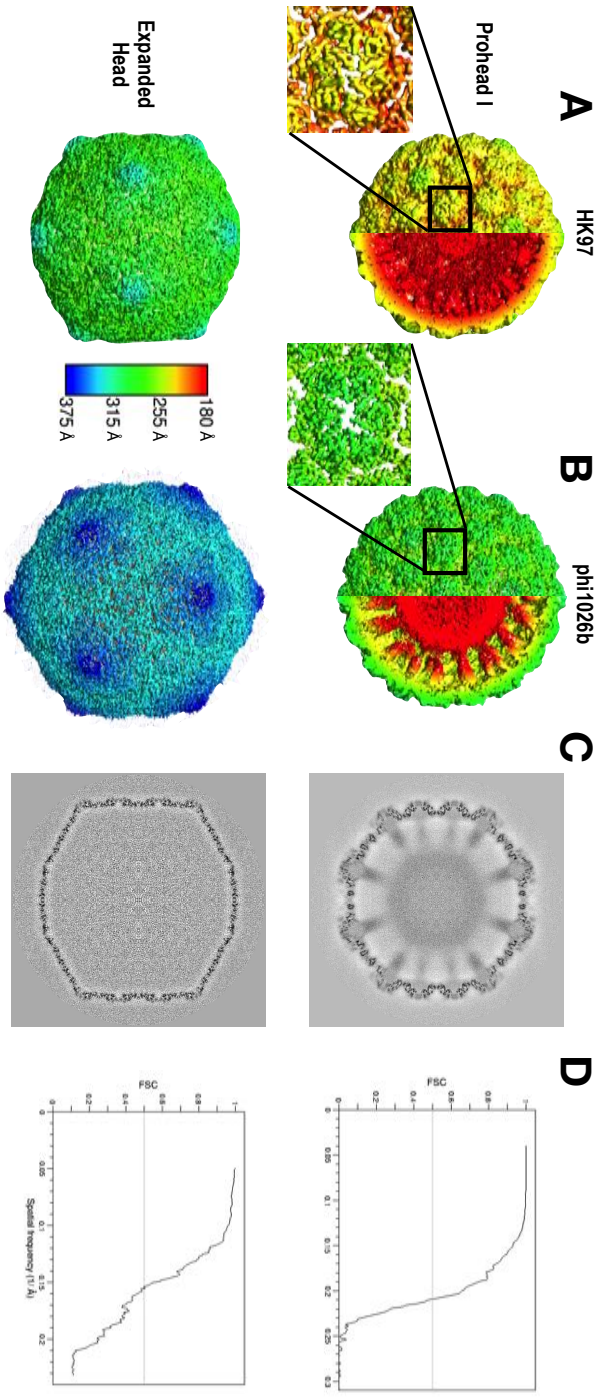

Figure 11. Reconstructions of Prohead I and Head capsids compared to the HK97 Prohead I and Head

A) Cryo-EM map of HK97 Prohead I generated by Dr James Conway. Expanded capsid of HK97 generated from EMAN2 pdb2map using PDB 1OHG at a resolution of ~5 Å. B) The Cryo-EM maps of Prohead I and Expanded Capsid for phi1026b. The interior density of Prohead I is shown. C) Sections of both phi1026b Prohead I and expanded capsid were made using trimpif from the Bsoft software package. D) FSC curves for both reconstructions show resolution.

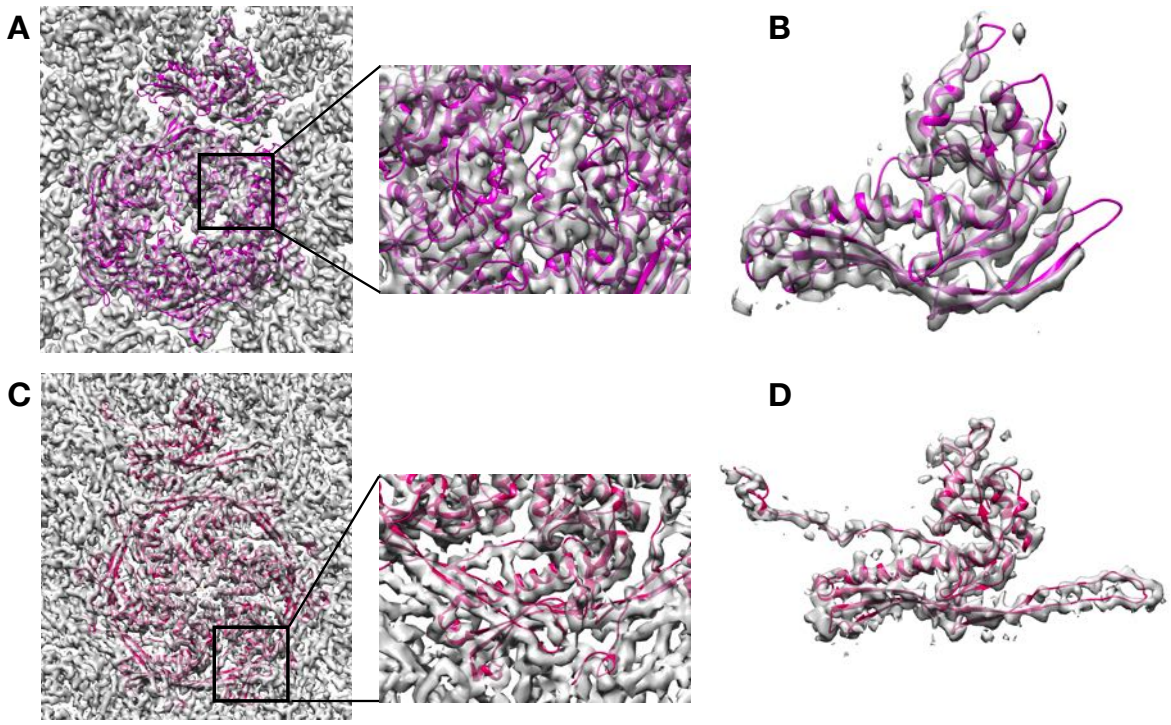

Figure 12. Rigid fitting of HK97 atomic models into phi1026b Prohead I and Head cryo-EM density

A) The atomic model of the asymmetric unit for HK97 Prohead I is rigid fitted into the phi1026b prohead density. The inset on the right highlights how the A domains of the HK97 model do not fit the density. B) A single monomer of the HK97 Prohead I model rigid fitted into the density of a single monomer. C) The atomic model of the asymmetric unit for HK97 expanded capsid is rigid fitted into the phi1026b expanded capsid density. The inset on the right highlights how the E loops and spine helices too not extend far enough to fit the phi1026b density. D) A single monomer of the HK97 expanded capsid atomic model rigid fitted into the phi1026b expanded capsid density of a single monomer. PDB-IDs: 3QPR, 1OHG

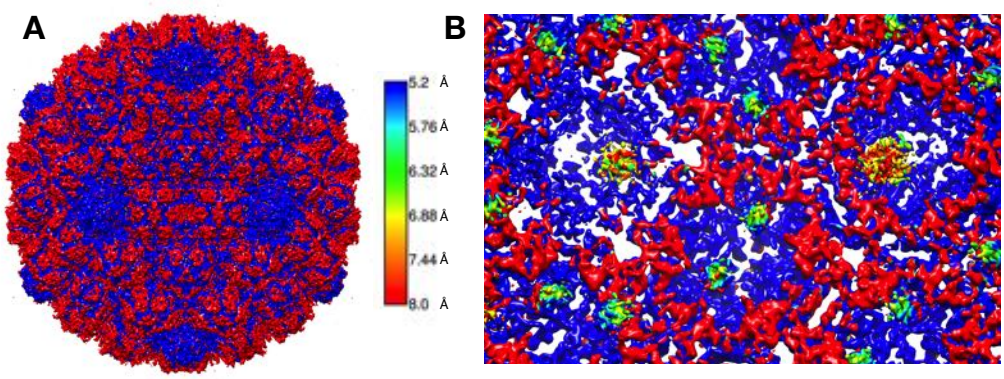

Figure 13. Resmap results for the phi1026b Prohead I structure

Resmap results showing local resolution of the A) external surface of the capsid and the B) interior of the penton and both hexons (Kucukelbir *et al.*, 2014).

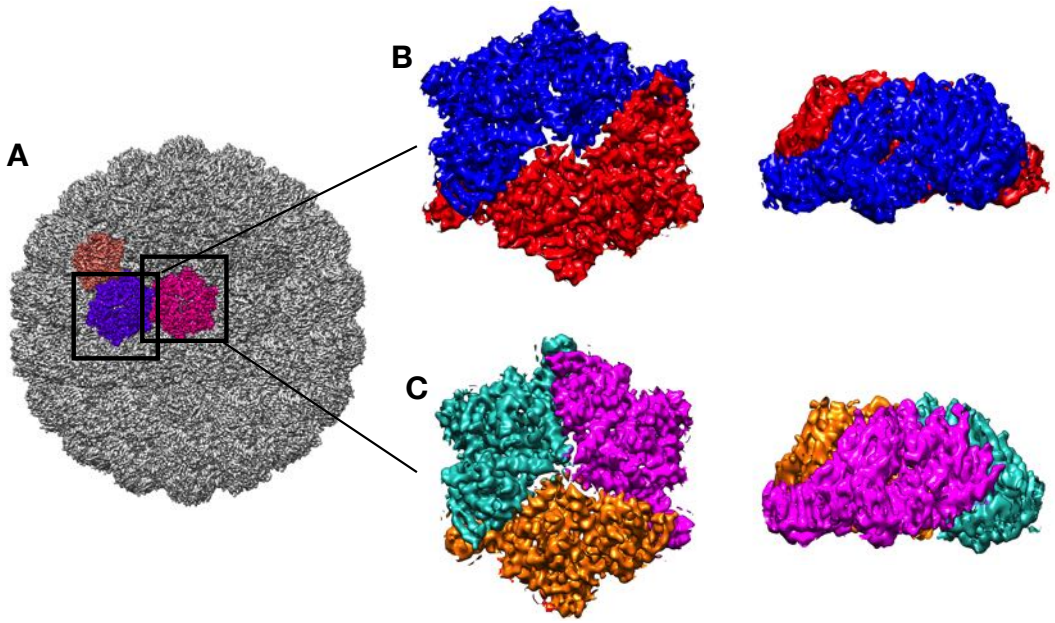

Figure 14. The hexons of the phi1026b Prohead I capsid

A) The phi1026b Prohead I capsid density is shown with the penton colored in salmon, the asymmetric hexon colored in purple, and the symmetric hexon colored in magenta. B) The asymmetric hexon is seen from above (left) and from the side (right). The skew is highlighted by coloring the montomers based on their placement in the skew. C) The symmetric hexon is shown from above (left) and from the side (right). The skew is highlighted by coloring the montomers based on their placement in the skew.

3.2.3 The phi1026b Prohead I capsid has highly organized interior density that reveal details of the delta domain's structure

The phi1026b prohead has large and distinct interior structures similar to those seen in HK97 Prohead I, but more well defined, suggesting they are less dynamic. The phi1026b Prohead I density map shows three regions with distinct characteristics: the capsid shell density, the interior columns and blobs of density under the hexons and pentons, and 'central mass' (Figure 15B). We interpret the interior columns and blobs of density as being major components of the scaffolding or delta domains, which are removed after assembly is complete. At high contour levels, the phi1026b delta domain columns appear to make a weak connection to the central mass, while the HK97 Prohead I delta domains do not appear to extend as far (Figure 15A). Given their locations in the prohead structure beneath hexameric, pentameric and pseudo-hexameric capsomers and the prior predictions that the delta domains should form alpha-helical coiled coils (Conway *et al.*, 1995), we attempted to model these parts of the structure with multistranded coiled-coil atomic models. A pentameric coiled-coil atomic model for a Tryptophan-zipper pentamer (Figure 15C)(PDB ID: 1T8Z)(Liu *et al.*, 2004) was found to fit into the columnar density under the pentons. A trimeric coiled coil atomic model for a mouse SynGAP (Figure 15D)(PDB ID: 5JXC)(Zeng *et al.*, 2016) was found to fit into the columnar density under the asymmetric hexons. Finally, a hexameric coiled coil atomic model for a synthetic protein (Figure 15E)(PDB ID: 3R46)(Zaccai *et al.*, 2011) was found to fit into the columnar density for the symmetric hexon. Cross sections of the columnar density with the coiled coil atomic models show that the density easily accommodates the models (Figure 16A-C). In Figure 15C-E, the density associated with the atomic models for the capsomers and the atomic coiled coils models are shown on the left to

illustrate that despite being considerably shorter than the 127 residues of the phi1026b delta domains, the ~80 residues long coiled coils models extend into the central mass density.

The relative thickness of the delta domain density suggests how the domains may be interacting. The delta domain density beneath the penton and the symmetric hexon is a single, columnar mass that we showed above can are reasonably interpreted as being composed of 5-stranded and 6-stranded coiled coils (Figure 15C,E). Spherical sections through the density map at radii 241Å and 238Å (Figure 17C), show dots of density that appear to correspond to the proposed coiled coils. The dots in the symmetric hexon delta domain density suggest a trimer of dimers arrangement that matches the orientation of the subunits in the capsomer. The density beneath the asymmetric hexon appear as two distinct columns that form an arch prior to making contact with the 'central mass' (Figure 15D). Each of the short columns beneath the asymmetric hexon are connected to interior shell of the prohead by three thinner strands suggesting a short 3-fold coiled-coil may form in this region (Figure 17A) and dots in the asymmetric hexon columnar density at radius 238 Å suggest such a trimer is formed (Figure 17C).

Hairpin-like structures on the interior of capsomers connect directly to the interior columns. In HK97, there are thick 'spokes' of density on the capsid interior that radiate from each of the local 3-fold intersections of capsomers and connect to the interior towers by strands of density, one strand per subunit (Duda, Conway, Oh, et al, unpublished). In phi1026b proheads, these spokes are replaced by distinct 'hairpin' like structures and make similar, one-strand-per-subunit connections to the interior columnar density (Figures 17A-C). The position of the hairpins and their connections suggests it is a combination of the N-arm of the mcp and the delta domain. The individual hairpins (9 different kinds) are located at slightly different radii, so they are not uniformly colored when radial-coded coloring is used, as in Figure 17A-C, and are not identical.

There is variation in shape and thickness in the surface-rendered views, suggesting that some hairpins are more dynamic than others. For example, the hairpins located under the asymmetric hexon oriented towards the 2-fold axis appear larger than the others (Figure 17A). The connections between the hairpin density and the columnar density are at ~90˚ angles and appear as "legs" that connect the columns or arches of the delta domain density to the capsomer density (Figure 18). These legs are well defined for all capsomer types and connections with some minor differences in density strength. These "legs" are not seen in the published HK97 density maps, but have been observed in newer, but unpublished HK97 Prohead I structures (Duda, Conway, Oh, et al, unpublished). These well-defined connections to the columnar density reveal that the asymmetric hexon trimeric delta domain density does not correspond to the asymmetric hexon skew axis (Figure 16D).

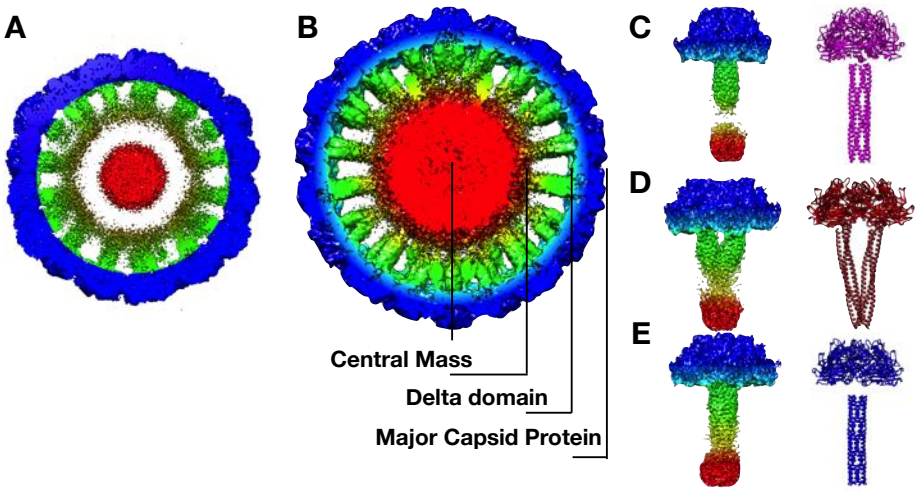

Figure 15. The delta domain density may extend into the central mass

A) A slice of the HK97 Prohead I density map showing the different regions of the density. This map was generated by the Conway lab for a previous study. B) A schematic detailing the different regions of density within the prohead. Colors are used to better delineate the regions of density. C) The penton delta domains appear to associate in one bundle which is fitted with several copies of a 5-fold coiled coil (PDB ID: 1T8Z). Left, the density associated with the atomic models for the penton and atomic models of coiled coils featured on the right. This is repeated for D and E. D) The asymmetric hexon has two distinct columns of density fitted with two 3-fold coiled coil (PDB ID: 5JXC). E) The symmetric hexon has one column of density which is fitted with several copies of a 6-fold coiled coil (PDB ID: 3R46)

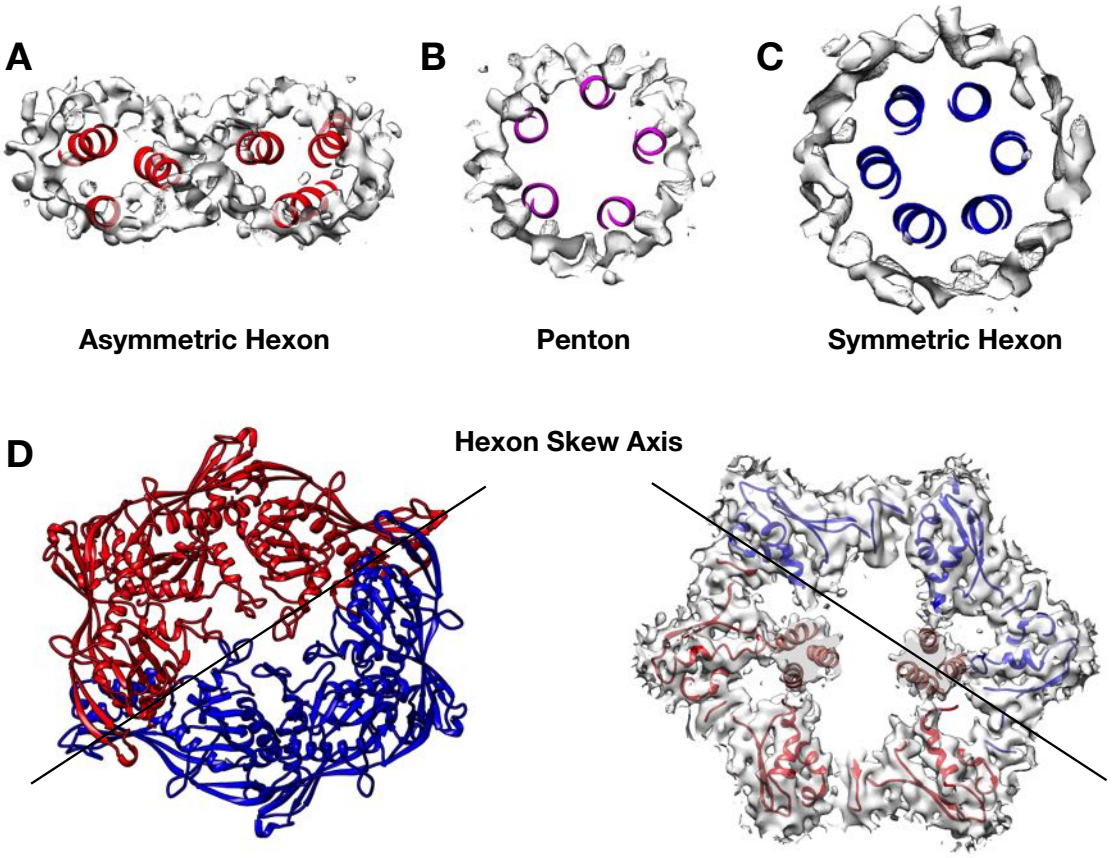

Figure 16. Atomic models of coiled coils fit the delta domain density well

A) A cross section of the asymmetric delta domain density is shown with coiled coils fitted (PDB ID: 5JXC). B) A cross section of the penton delta domain density is shown with coiled coils fitted (PDB ID: 1T8Z). C) A cross section of the symmetric hexon delta domain density with coiled coils fitted (PDB ID: 3R46). D) Top (left) and bottom (right) views of the asymmetric hexon are shown with the skew axis shown. On the right, coiled coils are included in the image (PDB ID: 5JXC).

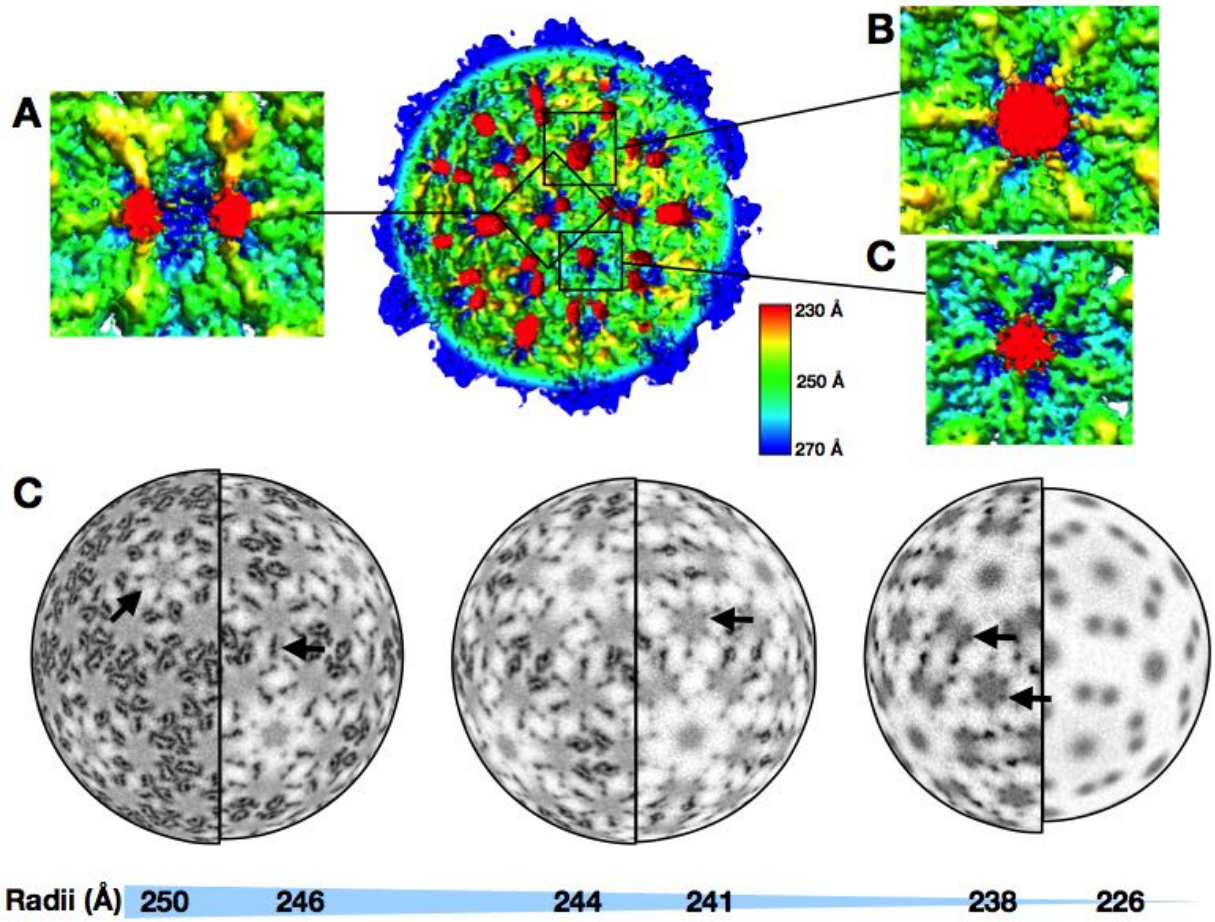

Figure 17. The internal spoke densities and the radical sections reveal features of the delta domains

A) The asymmetric hexon as viewed from the interior of the capsid shows three spokes connection to one column. B) The symmetric hexon has 6 spokes leading to one column. C) the penton has 5 spokes leading to one column. D) Radial sections of the capsid density at different radii show the connection between the spokes and the columns. At the radius of 238 Å and 241 Å it appears that there may be some relative strength of density suggesting helices.

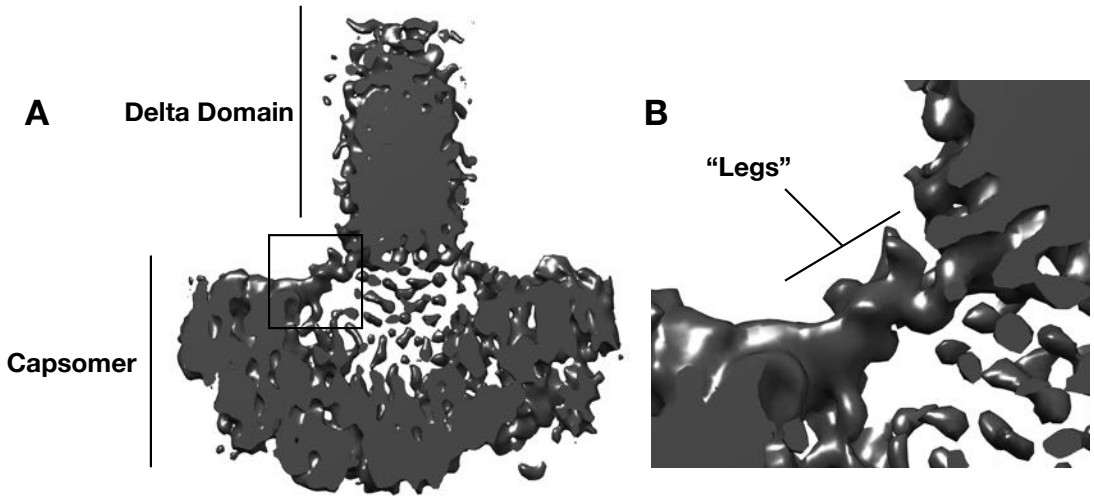

Figure 18. phi1026b Prohead I capsomers are connected to the delta domain density by well-defined density

A) A cross section of the symmetric hexon showing how the capsomer density and delta domain density is connected by two arches. B) A close up of one of the arches or 'legs' that connect the capsomer density to the delta domain density.

3.2.4 Modeling of the N-arm of the mcp

When rigid-body docking the HK97 Prohead I atomic model, there was additional density for which there was no corresponding atomic model. The hairpins are shown in Figure 19B, where the hairpin-shaped density, colored orange, is the density not occupied by the (HK97-based) phi1026b subunit models (that include only 5 residues of the N-arm and no residues of the delta domain). The hairpin-shaped density makes intercapsomer contacts at the trimeric interfaces. As shown in Figure 19B, there are different sized gaps at the trimer interfaces. These are apparent at an intermediate contour level that allows for identification of the HK97 fold domains but not significant differentiation of individual helices and sheets. The trimer interface that is formed near the pentons appear as the largest gaps. It is interesting to note that this interface includes the axes shown in Figure 21 (bottom row) that is most like those found in HK97 Prohead I. The smallest trimer interface gaps are found around the symmetric hexon. These gaps connect the interior

environment of the capsid to the external cellular host environment and are not present in the expanded capsid.

The connectivity of the hairpin-shaped density allows for some modeling. The hairpin-shaped density appears to primarily be associated with the internal surface of the prohead. From the N-terminal region of the E-loop the strand appears to make contact with the beta sheets that are located in the A-domain on the inner-facing surface of the capsomer. From there it then wraps back down under the P-domain. Here it appears to form a hairpin like structure that appears to be a combination of a beta strand, a coil, and a helix. Modeling was performed in COOT using a phi1026b prohead monomer that had already been fit to the density using MDFF. The final structures were then used for another round of MDFF to ensure better fitting into the density (Figure 20). The beta strand of the N-arm is aligned with the beta strands of the P-domain (Figure 20A) which aids with atomic modeling into this region of density (Figure 20B,C). Due to the lack of connectivity in the hairpin region, the density leading into the hairpin, as well as the lack of resolved bulky side chains, definitive atomic modeling of this density is not possible. On the other hand, it is possible to make an alpha carbon trace and it appears that the region is comprised primarily of the N-arm and some residues of the delta domain.

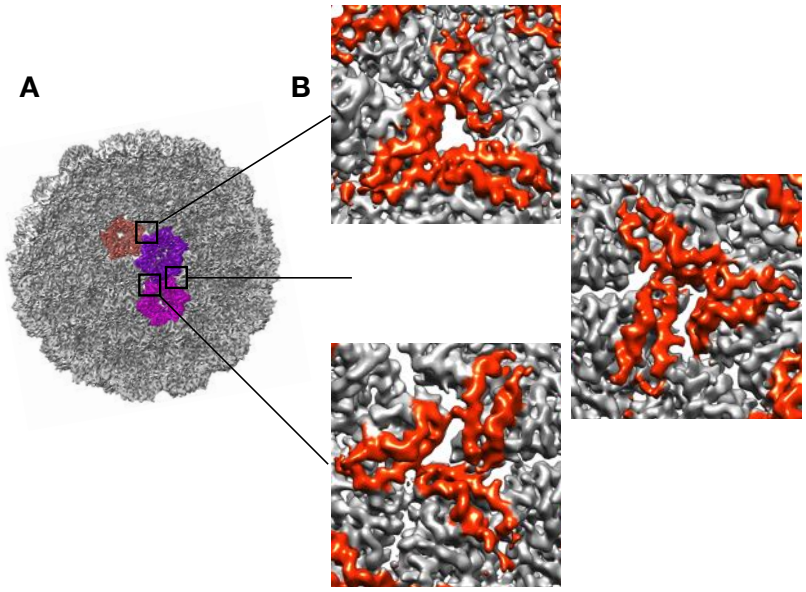

Figure 19. The N-arms on the interior surface of the Prohead I capsid are located at capsomer interfaces

A) An internal view of the Prohead I capsid with the penton, asymmetric hexon and symmetric hexons colored in salmon, purple, and magenta respectively. B) Hairpin density located at the capsomer interfaces highlighted in (A) are shown in closer detail with the density highlighted in orange.

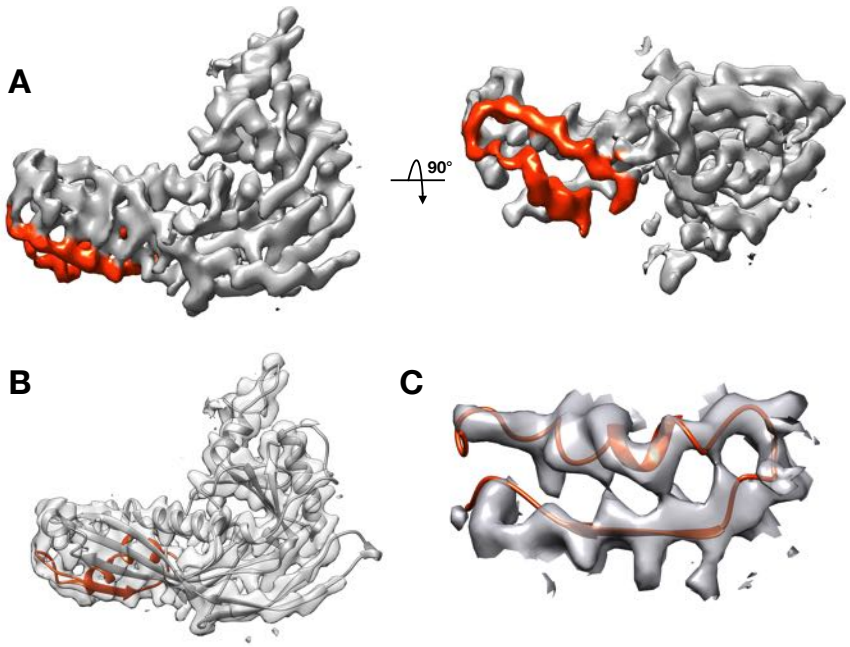

Figure 20. Modeling of the hairpin density

A) An individual monomer with the N-arm density colored in orange shown from the side (left) and from the bottom (right). B) The atomic model of the monomer shown inside the density with the N-arm colored in orange. C) The N-arm density with the atomic model colored in orange is shown.

3.2.5 The capsomers of phi1026b are larger than those of HK97

Both the asymmetric and symmetric hexons in the phi1026b Prohead I structure are larger than those in HK97 so measurements between conserved and similar residues in the two structures were taken (Figure 21). Measurements of distances in the hexon were taken from edge to edge and vertex to vertex. Distances between opposing vertices were made between F353, which is located in the P-loop, in the HK97 Prohead I and F384 in the phi1026b fitted Prohead I model. Distances between edges were made between D231, which is located in the G-loop, in HK97 and A262 in phi1026b. To measure the penton size, a measurement between vertices as well as a measurement between a vertex and an opposing edge were made. The pentons of the two phage were very similar in size with no more than 4 Å difference (Figure 21C, E). When analyzing the relative sizes of the hexons of the two viruses, it's interesting to note that the edge to edge and vertex to vertex axes that are most similar to HK97 are the same that make contact with the penton (Figure 21D, F). The edge to edge distance that is far longer than HK97 is the same that is in contact with the symmetric hexon which is not found in HK97. The vertex to vertex distance in the symmetric hexon is greater than that of both the HK97 and the phi1026b asymmetric hexon. It's important to note that the excess in size is not due to a loop that is not present in HK97. The size differences between the asymmetric hexons of the two viruses may be due to how the two sections of the hexon skew are oriented relative to each other or the way they interact. Despite the HK97 hexon being smaller in size than that of phi1026b, the A-domains at the skew interface are farther away from each other in HK97 than in phi1026b (Figure 22). This measurement was obtained by measuring the distances between the α carbons of conserved and overlapping residues in the A domains of HK97 (D256 and I285) and phi1026b (D287 and I316). Further, the angles of the monomer interfaces are greater in HK97 than in phi1026b. This measurement was obtained by measuring the angle between planes

formed by the α carbons of the residues 210, 267, and 284 in HK97 and the α carbons of the residues 241, 298, and 315 in phi1026b. Together, this suggests that the skewing of HK97 asymmetric has greater 'lifting' than 'shifting' at the skew interface than that of phi1026b.

In the Head, both the pentons and hexons in phi1026b are larger than those in HK97. Figure 23 shows distance measurements between features in the two structures. Distances between vertices were made between V354 in the HK97 expanded capsid crystal structure (PDB ID: 1OHG) and E385 in the phi1026b fitted expanded capsid structure. For the penton, a measurement between the P-loop vertex and the center of the opposing 3-stranded β sheets was made (A147 for HK97 and S178 for phi1026b). The phi1026b penton is 3 Å larger along both axes measured (Figure 23C, E). The measurements for the penton adjacent hexons in phi1026b shows that the axis that makes contact with the extra hexon is nearly 7 Å longer than that in HK97 (Figure 23D,F). This is the axis that shows the most difference between the two viruses.

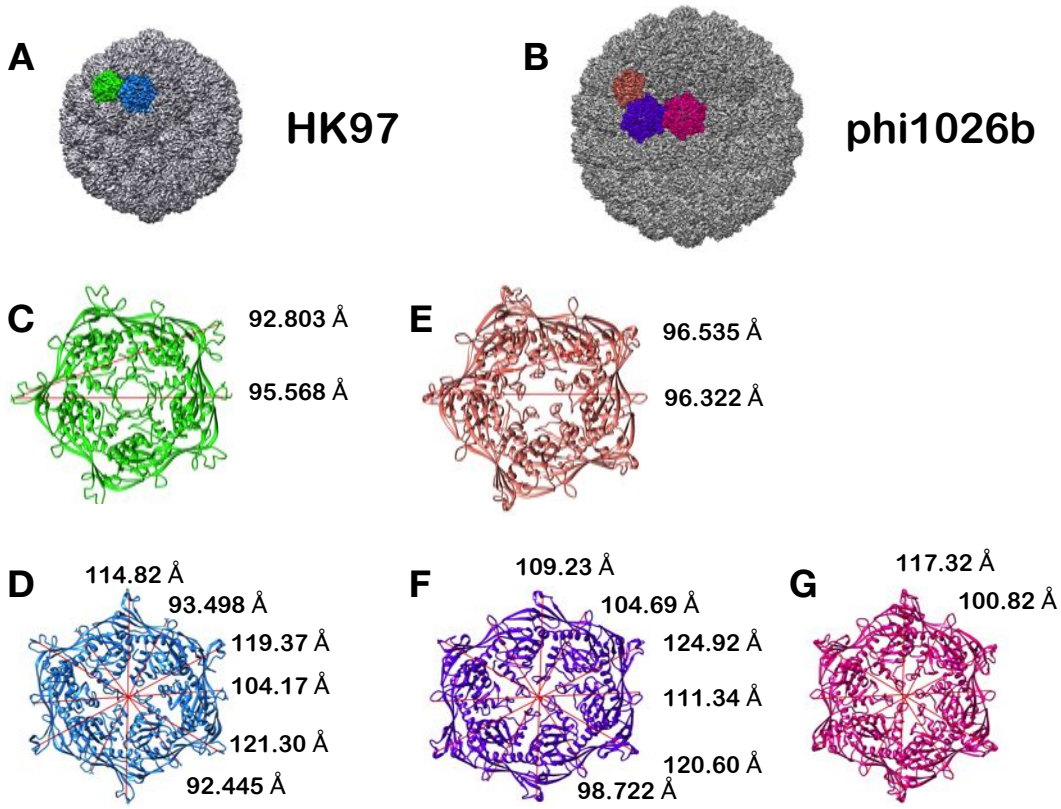

Figure 21. The phi1026b Prohead I capsomers are larger in some axes compared to HK97 Prohead I

A) The HK97 Prohead I structure is shown with the penton and hexon highlighted in green and blue respectively. B) The phi1026b Prohead I structure is shown with the penton, asymmetric hexon, and symmetric hexon highlighted in salmon, purple, and magenta respectively. Atomic models of the capsomers are shown below with measurements located at the endpoints of the measured axes. C, D) The atomic models for the HK97 Prohead I penton and hexon. E,F,G) The atomic models for the phi1026b Prohead I penton, asymmetric hexon, and symmetric hexon.

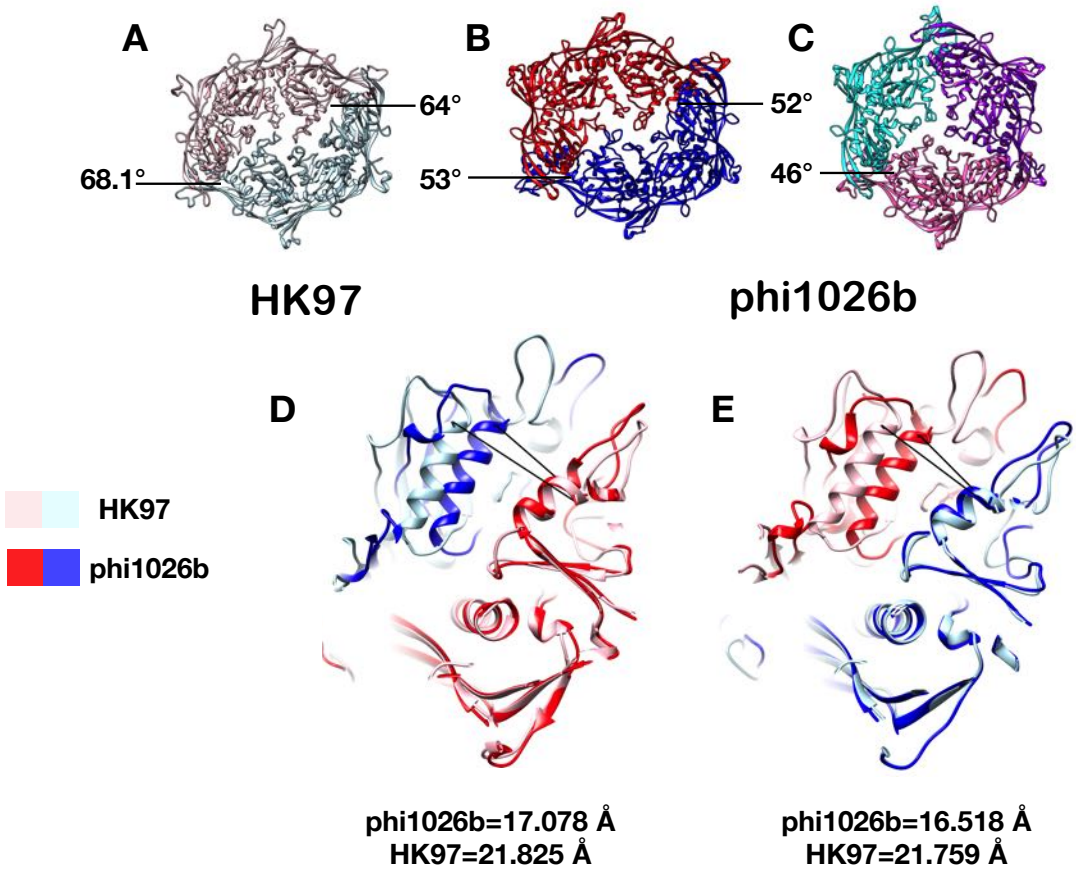

Figure 22. The skewing of the asymmetric hexon in HK97 and phi1026b examined geometrically

A) The HK97 hexon is shown, B) the phi1026b asymmetric hexon, and C) the phi1026b symmetric hexon are shown with the skew interface angles shown. Angles were measured in the program Chimera by measuring the angles between planes generated from the α carbons of the residues 210, 267, and 284 in HK97 and the α carbons of the residues 241, 298, and 315 in phi1026b. D,E) The distances between the α carbons of conserved and overlapping residues in the A domains of HK97 (D256 and I285) and phi1026b (D287 and I316). HK97 is colored in pink and light blue. Phi1026b is colored in red and blue.

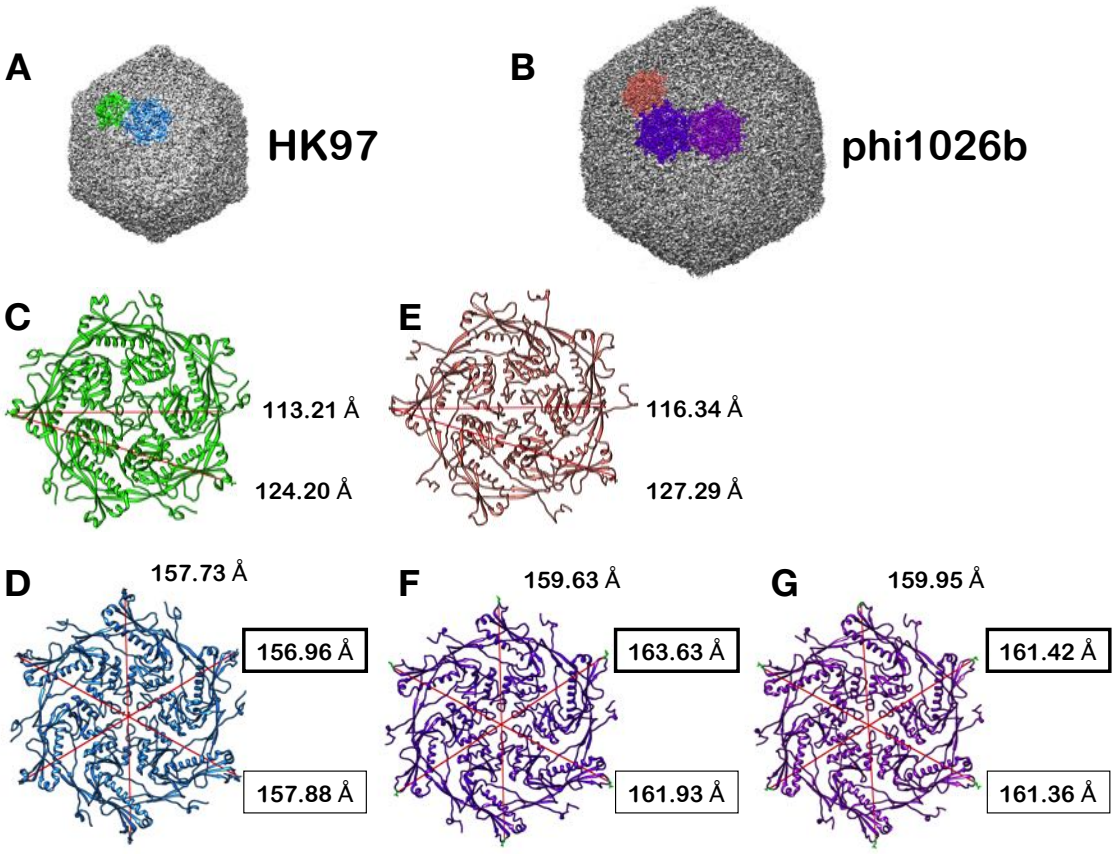

Figure 23. The phi1026b expanded capsid capsomers are larger in some axes compared to HK97

A) The HK97 Head structure is shown with the penton and hexon highlighted in green and blue respectively. B) The phi1026b expanded capsid structure is shown with the penton, asymmetric hexon, and symmetric hexon highlighted in salmon, purple, and magenta respectively. Atomic models of the capsomers are shown below with measurements located at the endpoints of the measured axes. C, D) The atomic models for the HK97 Head penton and hexon. E,F,G) The atomic models for the phi1026b expanded capsid penton, asymmetric hexon, and symmetric hexon.

3.2.6 Geometric analysis of intercapsomer contacts showed differences between HK97 and phi1026b

In prohead I, the angles at the dimer interfaces are different between HK97 and phi1026b (Figure 24). Angles were measured by selecting the alpha carbons of three residues (HK97: A192, F176, T380) (phi1026b: A223, F207, A411), generating a plane connecting the three atoms, and measuring the angles between the planes. These residues were chosen based on conservation or location. Alignments of the HK97 and phi1026b dimer interface models were based on one

monomer in the dimer. After alignment these three residues were overlapped suggesting they would be a reliable for the angle measurements at the dimer interface. With exception of the penton-hexon interface, all dimer interface angles are larger in phi1026b compared to those of HK97 (Figure 24). In the T=7 icosahedral lattice, there are 4 unique dimer interfaces and in the T=9 icosahedral lattice, there are 5 unique dimer interfaces (Dr Robert Duda, personal communication). In HK97, the blue interface is comprised of chains E and D. This pair is replicated adjacent to this interface but is the mirror image with chains D and E. In phi1026b, the unique interface sits at the mirror image with chains D and D. Since it retains chain D, it is possible to compare it to the interface with monomer comprised of D and E in HK97. While the interface between chains E and D in HK97 and phi1026b differ by ~11 degrees, the interface with chains D and D differs from HK97 by ~18 degrees. Compared to the same interface in phi1026b it differs by ~8 degrees. While it seems intuitive that there would be differences of angles between the dimer interfaces based due to the difference in curvature between the two viruses, it is interesting to note that the degree of difference is not the same for all interfaces as the difference ranges from ~7 to ~20 degrees.

In the expanded capsid, the angles at the dimer interfaces are different between HK97 and phi1026b (Figure 25). Angles at the dimer interfaces were measured using the same residues and method as previously described for the Prohead I structure. The dimer interface located at the penton have angles that are the most similar between the two viruses. The angle differences at the interfaces are not the same between the Prohead I structure and the expanded capsids. For example, the green dimer in Figures 24 and 25 is different by ~7 degrees in the Prohead I structure with the HK97 dimer being less than that of phi1026b but in the expanded capsid there is a ~7 degree difference between the dimers with the phi1026b dimer being less than that of the HK97 dimer.

This relationship is the same for the blue and orange dimer interfaces in Figures 24 and 25. Interestingly, the dimer interface unique to the T=9 icosahedral lattice (colored in magenta in Figures 24 and 25) does not undergo a large change between Prohead I and expanded capsid structures.

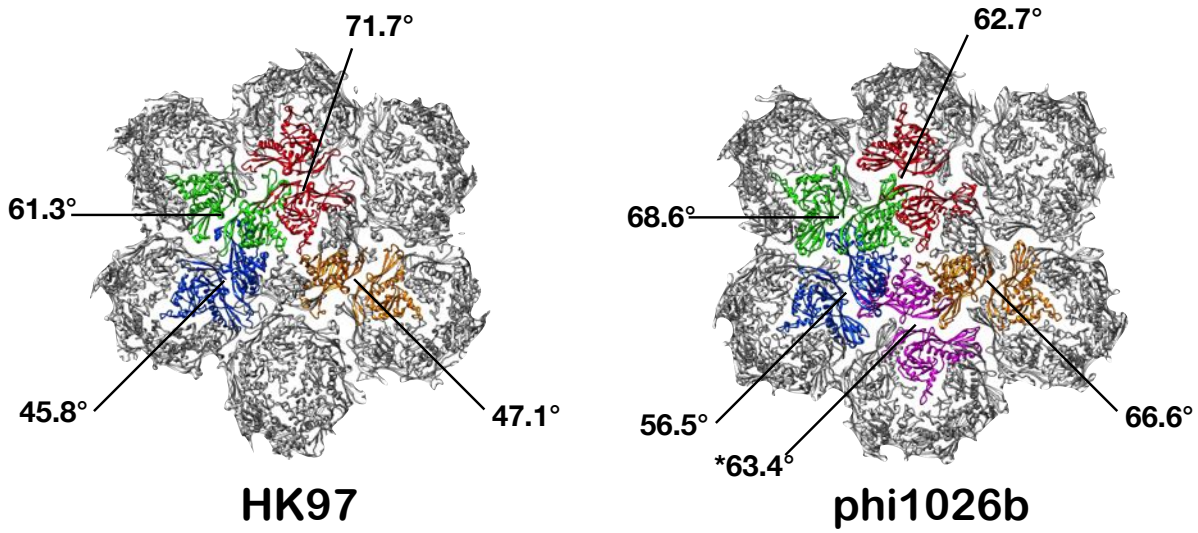

Figure 24. Angles at the dimer interfaces of HK97 and phi1026b Prohead I

Individual dimers are represented by different colors. In HK97, the dimers are: chains G and A in red; chains F and B in green; chains E and D in blue; and chains C and C in orange. In phi1026b, the dimers are: chains G and A in red; chains F and B in green; chains E and H in blue; chains C and I in orange; and chains D and D in magenta.

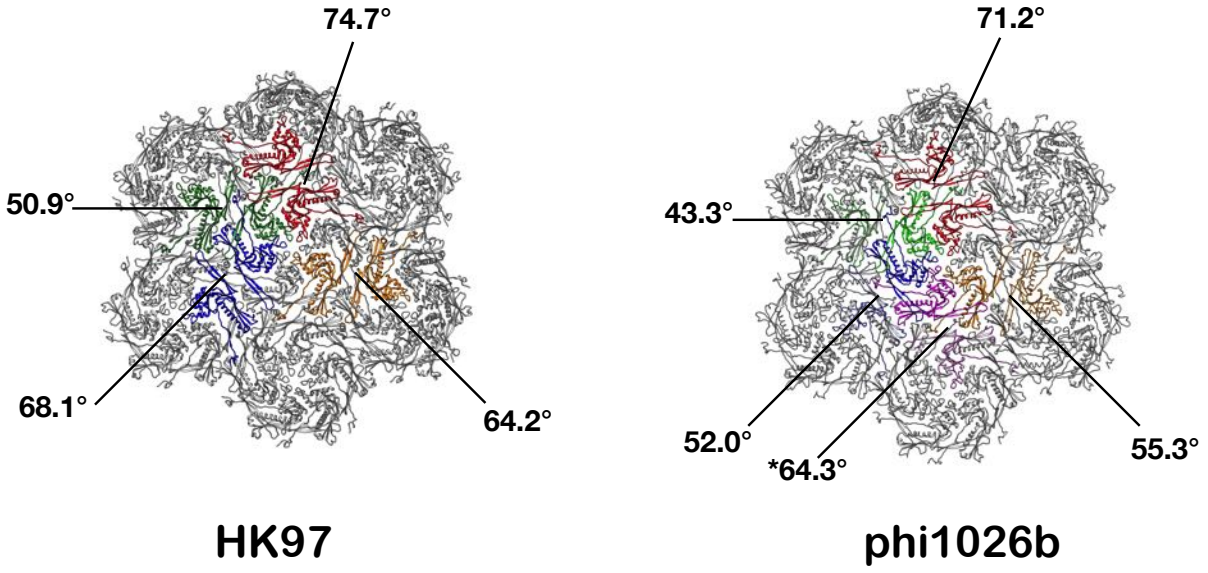

Figure 25. Angles at the dimer interfaces of HK97 and phi1026b expanded capsid

Individual dimers are represented by different colors. In HK97, the dimers are: chains G and A in red; chains F and B in green; chains E and D in blue; and chains C and C in orange. In phi1026b, the dimers are: chains G and A in red; chains F and B in green; chains E and H in blue; chains C and I in orange; and chains D and D in magenta.

3.2.7 The capsomers interfaces in phi1026b are different than those in HK97

During HK97 assembly, salt bridges are made between the residues D231 and K178 which are located on the G-loop and E-loop respectively. These form bridge-like densities between capsomers in prohead reconstructions (Figure 26) that result in a 'high radius' contact between the capsomers. These residues can be substituted with hydrophobic residues of similar size and assembly can still proceed normally (Tso *et al.*, 2014). In phi1026b, residues of similar size or type are not conserved at these sites and bridge-like densities are not found between the G-loop and E-loops densities (Figure 26). While such connections are not found in phi1026b, high radius contacts are found between the G-loop and P-loop at the interface between the symmetric and asymmetric hexons (Figure 27). A key difference about these high radius contacts is that in HK97

they occur between two monomers but in phi1026b they involve four separate monomers. While these bridge-like densities are fairly distinct in the phi1026b density, a specific pair of residues that constitute the pairing are not clear. This might be due to a hydrogen bond being formed between one of the charged residues on the P loop with the backbone of the G-loop. No other high-radius densities were found between the capsomers in phi1026b.

In HK97 the residues E344, E348, E363, R194, and R347 are in contact at the 3-fold axis throughout maturation playing a 'staple' like role (Gertsman *et al.*, 2009). As shown in Figure 9C, the HK97 residues R194, E348, E363 are conserved as R225, E379, and E394 in phi1026b. The HK97 residues E344 and R347 are converted to L375 and T378 in phi1026b. It's important to note that the 3-fold axis in phi1026b is in the center of the 'symmetric' hexon for Prohead I or the 3-fold axis hexon in the expanded capsid. On the other hand, there are 3 positions where the P domains form a trimer interface in phi1026b Prohead I and expanded capsid, which are shown in Figure 27. In Prohead I, residues 225 and 394 appear to be in contact in at least two of the three unique trimer interfaces. The rest are not near enough to each other to make contact. They do not make any other unique pairs with residues on neighboring monomers. Further, they appear to have strong intramonomer contacts as evidenced by the ring-like protrusions at some interfaces (Figure 28A, box). In the expanded capsid, 225 and 394 are consistently in contact at all the trimeric interfaces. What made the 'staple' residues unique is that the contacts were consistent in both prohead and the expanded capsid yet that does not appear to be the case here.

In the expanded capsid, there appears to be extra density between the capsomers. This is most likely the additional residues that are present in the phi1026b mature capsid domain. When the HK97 expanded capsid model is fitted into the expanded capsid density for phi1026b, there are two distinct strands of density that are unoccupied (Figure 29). The N-arms of the HK97 model

are not long enough to fit this density after flexible modeling. Since it has been hard to establish the extent of cleavage of the phi1026b mcp, it's not clear whether this extra density is biologically relevant. It is possible that if the delta domains are not cleaved that somehow the delta domain rearranged to the external surface of the capsid. While it seems unlikely, it cannot be ruled out at this time. Despite the addition of the extra density at the dimer interface, the distance between capsomers in the expanded capsids is minimally different between HK97 and phi1026b. Alignment of mcp amino acid sequences for HK97 and phi1026b reveal that there are 11 additional residues in the phi1026b mature capsid domain. 9 of those are in the N-terminal region. It is very likely that these additional strands of density between the capsomers are simply the slightly longer N-arms of the phi1026b mcp.

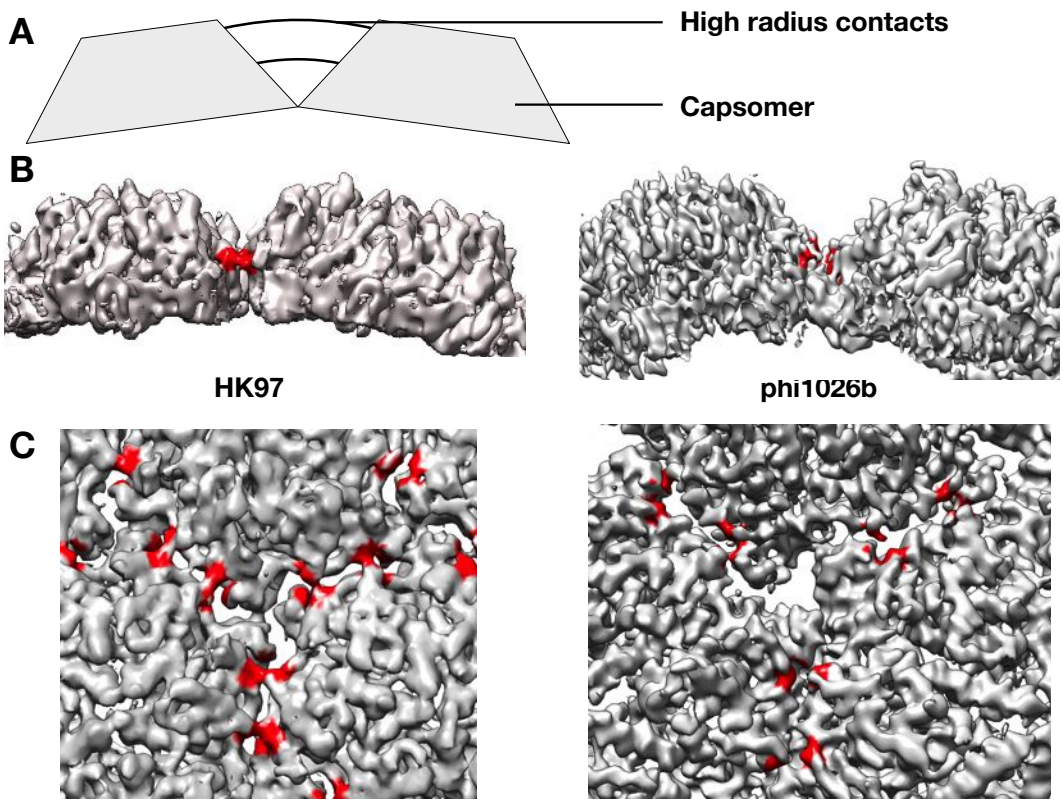

Figure 26. The high radius intercapsomer contacts found in HK97 are not conserved in phi1026b

A) A schematic showing how what 'high radius' contacts are relative to the hill-like capsomers. B) On the left, the intracapsomer contacts of D231 and K178 are shown in red are shown from the side. On the right, the equivalent residues (A262, L209) in phi1026b are shown in red. C) On the left, a top view of the intracapsomer contacts formed by D231 and K178 in HK97 are shown in red. The contacts form a bridge between capsomers. On the left, the equivalent residues in phi1026b, G261 and N208, are are shown in red and do not form bridges between the capsomers.

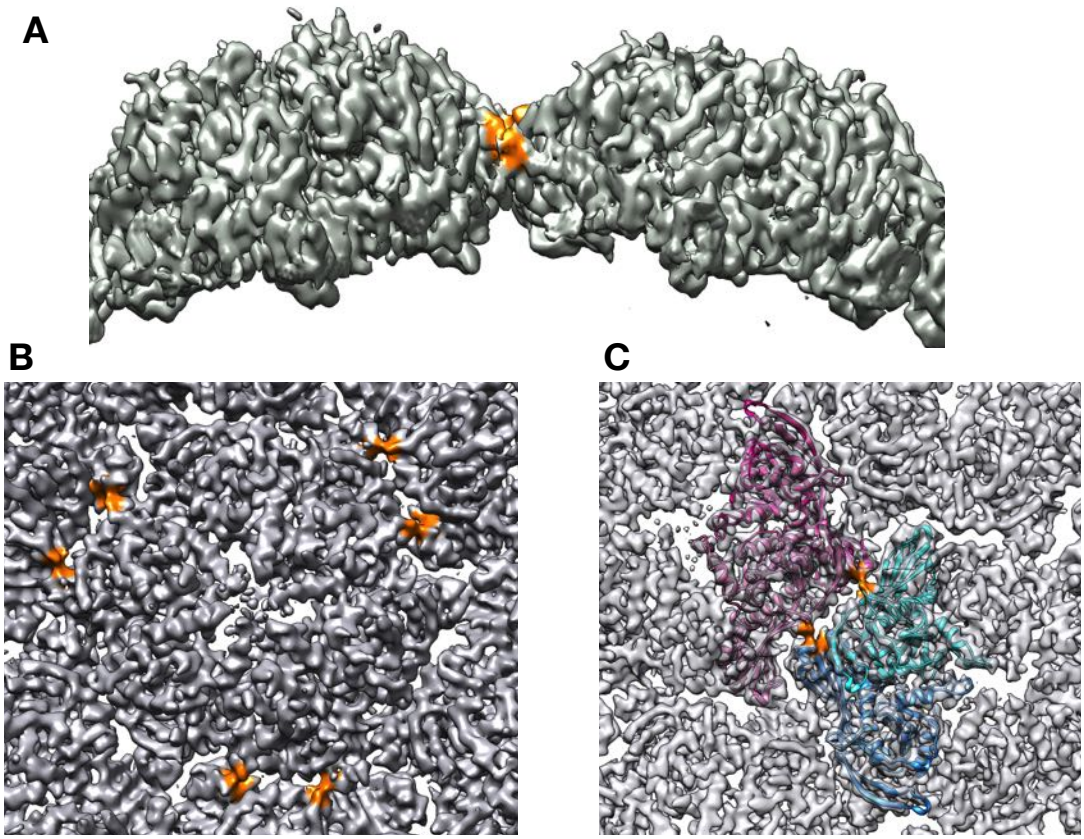

Figure 27. A high radius contact is found between the symmetric and asymmetric hexons

A) The high radius contact formed between the G loop and the P loop is shown in orange. B) A top view of the bridges formed at a high radius are shown in orange. C) The high radius contacts involve four different monomers which are shown in the density with the bridges colored orange.

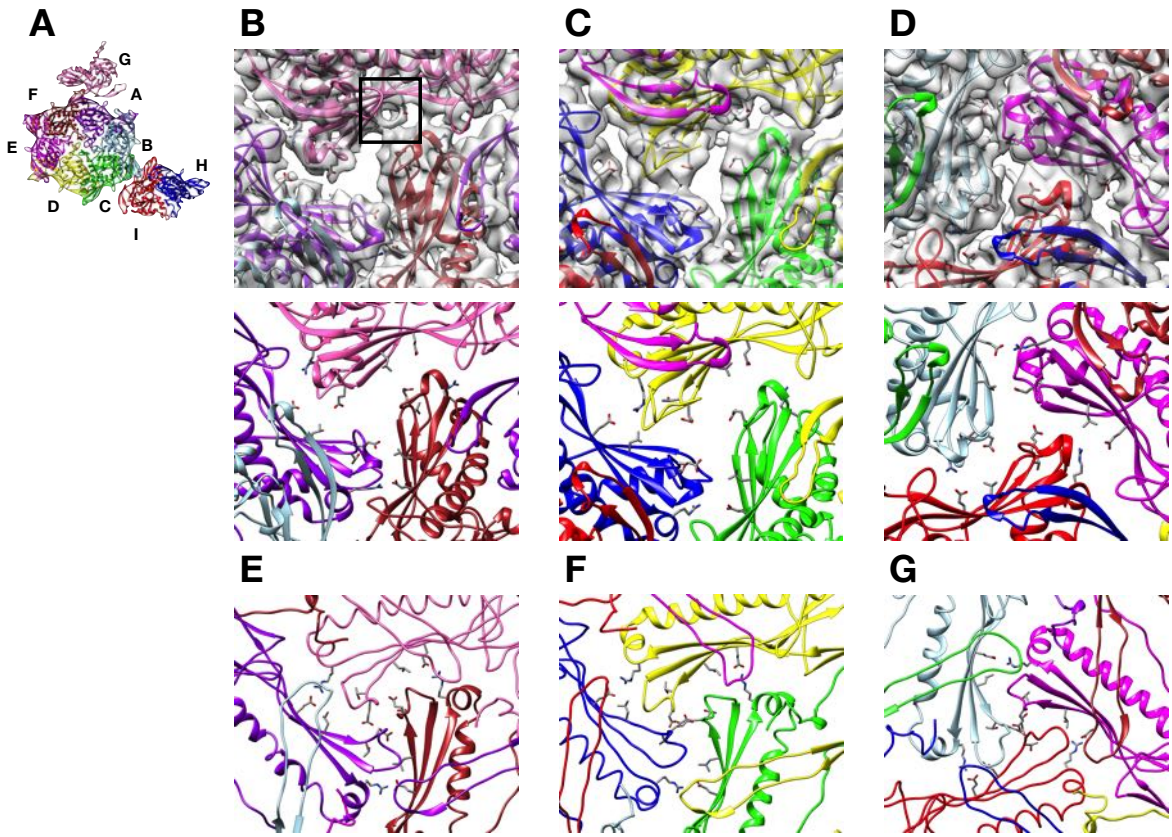

Figure 28. Trimeric interfaces in the phi1026b Prohead I and expanded capsid

A) The phi1026b asymmetric unit is shown with the chains in the asymmetric unit labeled. A-C) The trimeric interfaces of the Prohead I are shown with (top row) and without (bottom row) the corresponding density. D-F) The trimeric interfaces of the expanded capsid are shown.

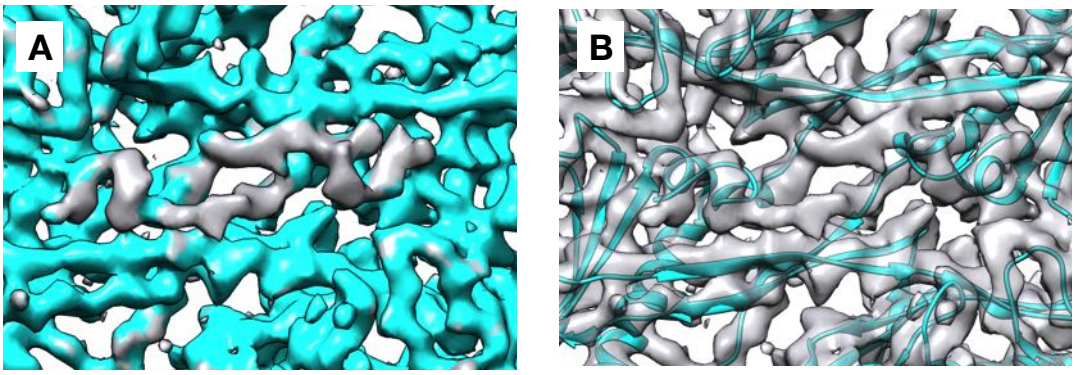

Figure 29. There is extra density at the capsomer interfaces in the phi1026b capsid

A) The density that is occupied when monomers of HK97 PDB model 1OHG are rigid fitted into the density is colored teal leaving some density uncolored. B) The same density showing the HK97 expanded capsid model in the density. In the center of the image, there are two distinct strands that are unoccupied by model.

3.2.8 Differences in intracapsomer contacts between HK97 and phi1026b

The E-loop to spine helix ionic interaction between E153 on the E-loop and R210 on the spine helix is critical to assembly (Hasek *et al.*, 2017). Further work in the lab identified that HK97 residues E269 and R372 participate in an ionic network with 153 and 210 (Dr Duda, personal communication). These residues are conserved in phi1026b as E184, R241, E300, and R403. Comparing, HK97 Prohead II PDB 3E8K, the orientation of these 4 residues is similar in HK97 to phi1026b in most cases. The density associated with E184 and R241 suggest a salt bridge in all cases but one case (Figure 30). In chains D and C, the orientation of the residues is different, suggesting that the connection between the E-loop and spine helix is broken in this case. Note that this alternate orientation also occurs along the hexon axis that is longer than that found in HK97. There were three different types of orientations found between all the monomers (Figure 30D-F). The connection shown at the top of the image that is boxed is the most similar to the arrangement found in Prohead II of HK97.

The wedge-shaped A domains of HK97 fill the central (axial) regions of the capsomers. In Figure 31, the A domain helices are highlighted in red and orange in the penton adjacent hexon for both the Prohead I structure and the expanded capsid. In the Prohead I structure, 4 of the A domain helices are nearly parallel with exception of those at the skew interface while in the expanded capsid all of the alpha helices of adjoining monomers are almost completely parallel. HK97 has 11 charged residues that interact distributed evenly along the A-domain interface. In phi1026b, there are 4 charged residues (D287, R290, E317, K320) at this interface and they are clustered towards the top or center of the capsomer. When looking at the locations and types of connections at the top of the Prohead I A domains, a pattern of interactions appears (Figure 32). Both charged and uncharged connections are not made at the skew interfaces in the asymmetric and symmetric hexons that are made at all other monomer interfaces (Figure 33).

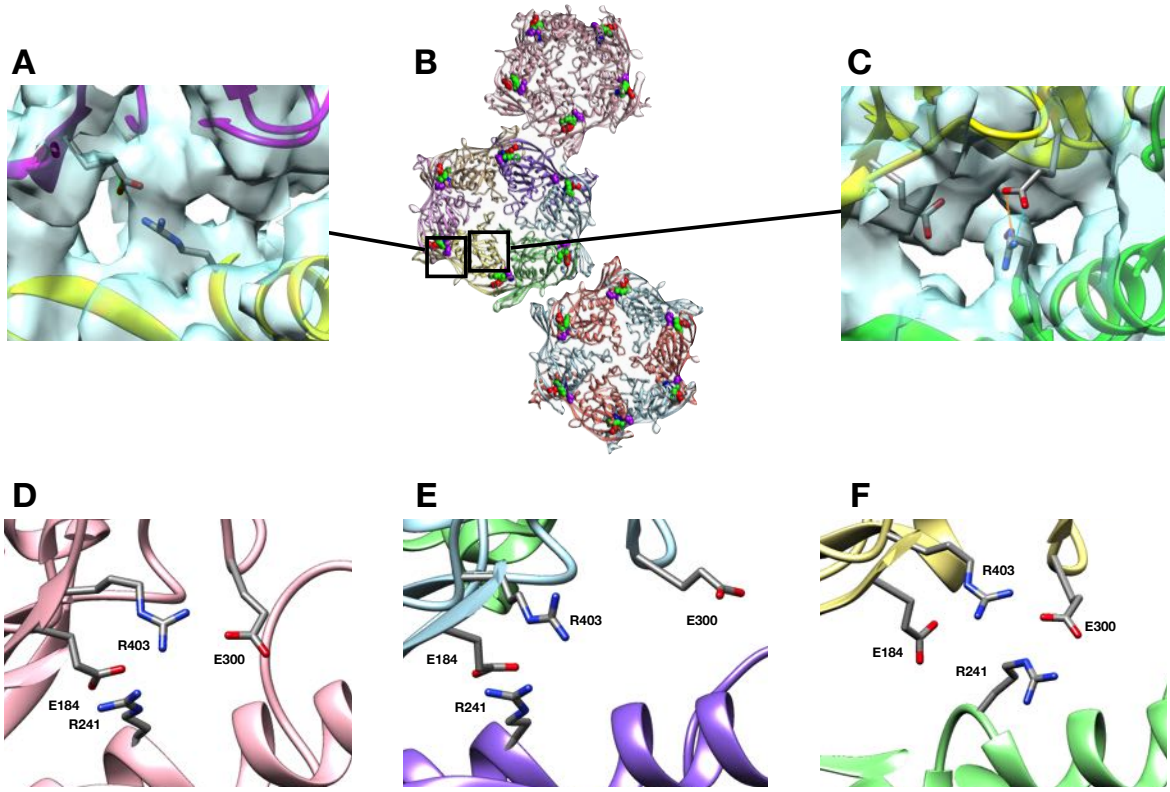

Figure 30. The E-loop to spine helix salt bridge in the asymmetric hexon is different

A) The conserved connection between E184 and R241 between chains E and D (magenta and yellow) is shown with the accompanying density. B) The atomic models for the penton, asymmetric hexon and symmetric hexon are shown with E184, R241, E300, and R403 highlighted. C) The broken connection between the E-loop and Spine helix between chains D and C (yellow and green) is shown in the atomic model and accompanying density. D,E,F) Three types of arrangements of residues E184, R241, E300, and R403 were found in the asymmetric unit and representatives of each are shown.

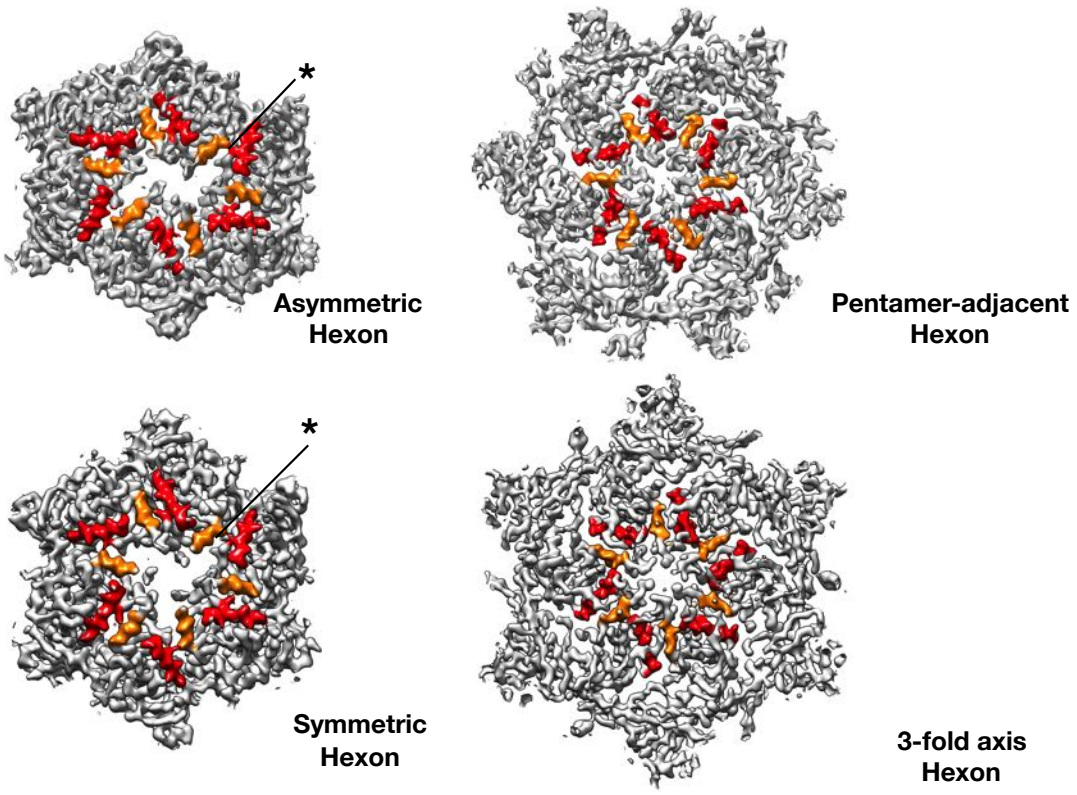

Figure 31. The orientation of the A-domain helices in phi1026b Prohead I and phi1026b Head hexons

Skewed interfaces are marked with an asterisk

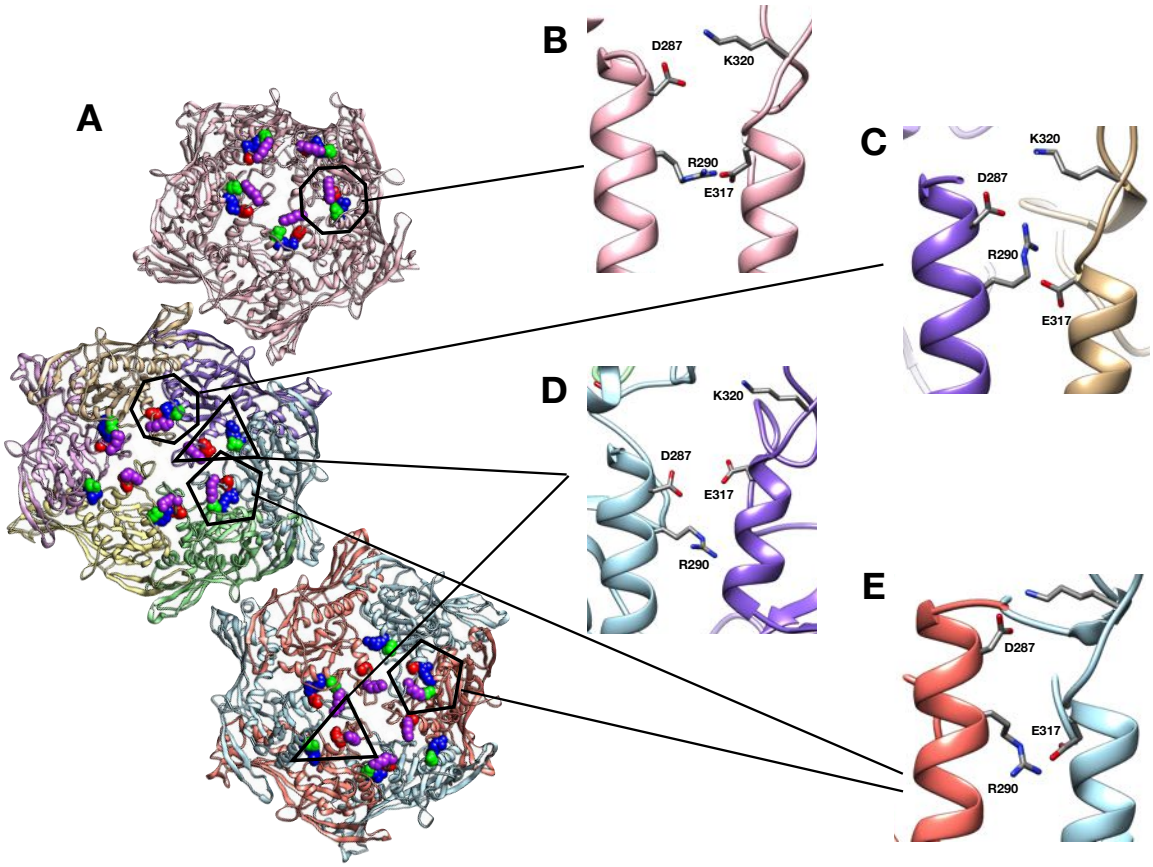

Figure 32. The connections found in the A domains of the phi1026b Prohead I capsomers

A) The three types of capsomers found in phi1026b with several residues highlighted: D287 (E. F. Pettersen *et al.*), R290 (blue), E317 (red), K320 (purple). B,C,D,E) Four different arrangements of the residues were found throughout the capsomers and they are displayed on the right with lines pointing to where they are found.

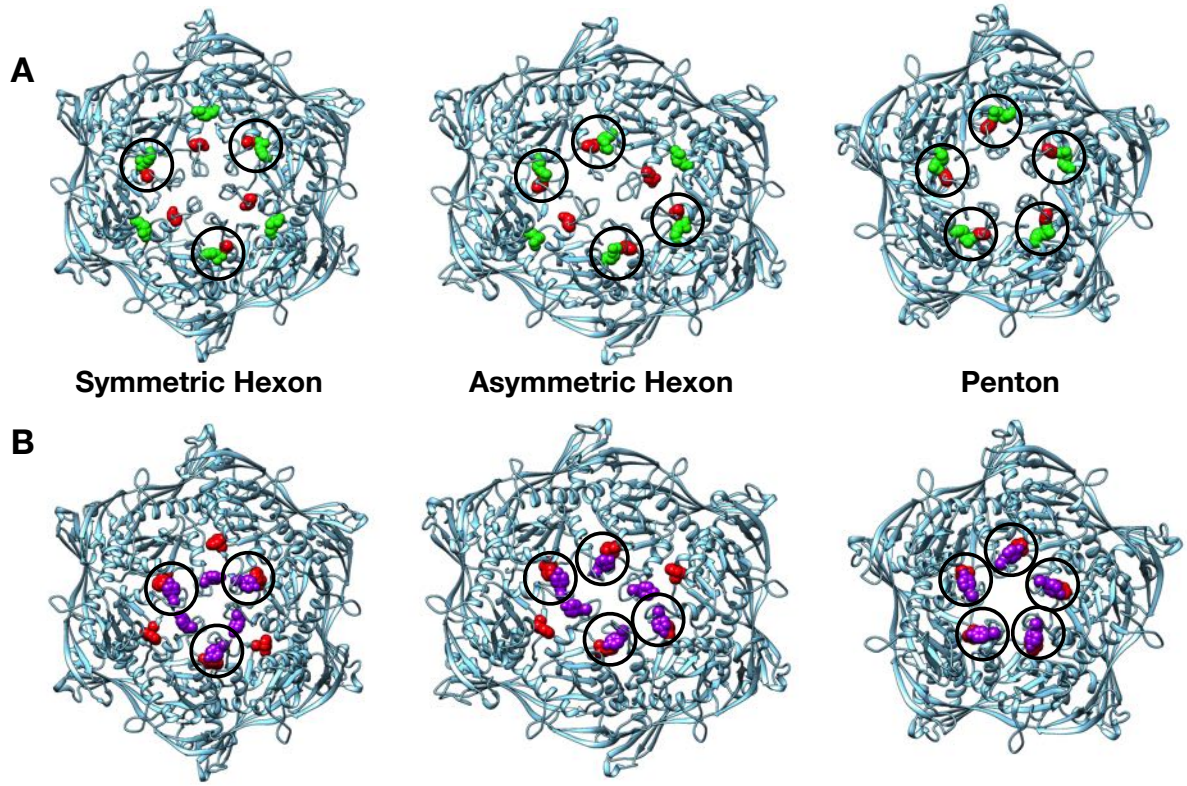

Figure 33. Alternate interactions at the top of the A-domain in the phi1026b Prohead I capsomers

A) The connections (and lack thereof) between residues R290 (E. F. Pettersen *et al.*) and E317 (red) found near the top of the A-domains. B) The same is shown for residues Y326 (purple) and W340 (red).

3.3 Discussion

Our phi1026b expression system made Prohead I particles instead of Prohead II

We expressed the genes for the mcp and the protease, but only Prohead I and expanded capsid particles were found in the final preparation. The majority of the particles were Prohead I capsids with no more than ~5000 expanded capsid particles. The central density of the phi1026b Prohead I structure may be the protease (Veesler *et al.*, 2014), proteolysis products, accidentally incorporated host proteins, and noise. The presence of an additional mcp protein band that is missing ~13 amino acid residues suggests that the proteolysis of the delta domain proceeds in

steps. The presence of a ~31 kDa protein band in fraction samples of early purification steps suggests a small population of Prohead II intermediates, where the delta domain is proteolytically removed but the capsid has not yet expanded, were generated. However, the lack of Prohead II intermediates in the later steps of purification suggests that the phi1026b Prohead II structure is not stable and rapidly expands to the mature confirmation. This is unlike Hk97 where both the Prohead I and Prohead II assembly intermediates are stable enough for purification and analysis (Duda, Hempel*, et al.*, 1995). The small number of expanded capsids compared to the Prohead I capsids as well as the partial cleavage of the mcp suggests that the protease is relatively ineffective. The lack of a ~31 kDa mcp protein band as well as higher molecular weight bands in the final purification step in addition to the population of expanded capsids suggests that once the mcp is fully cleaved and the capsid undergoes expansion, the crosslinking reaction is highly efficient.

This work shows that phi1026b is capable of assembling into regular capsids in absence of the portal. This feature is shared with HK97 but not all phage that utilize the HK97 fold and delta domain can assemble into regular structures without the portal (Huet, Duda*, et al.*, 2016). In the case of T5, portal is required for capsid size determination. When the genes for the T5 portal were not expressed with the capsid protein, tubes as well as irregular particles were isolated (Huet, Duda*, et al.*, 2016). It was posited that the portal may be required for larger capsid assembly as there are nearly twice as many mcp monomers than HK97 required to form the T5 capsid. While phi1026b is comprised of more mcp subunits (535 monomers) than HK97 (415 monomers), the portal is not needed for its assembly.

The symmetric hexon of the phi1026b Prohead I capsid may provide new insights.

Thus far, 'symmetric' (as opposed to a skewed) hexons have been seen in 2 prohead capsids that feature the HK97 fold: the T=12 marine SIO-2 (*Siphoviridae*) (Lander *et al.*, 2012) and the T=3, Q=5 phi29 (*Podoviridae*) (Morais *et al.*, 2005). In both viruses, the 'symmetric' hexons sit on the 3-fold axis (Figure 34). The relative size of the SIO-2 symmetric hexon led the authors to suggest that it is an intermediate state between the skewed prohead hexon and the mature capsid hexon (Lander *et al.*, 2012). In phi29, skewed hexons are only present on quasi-symmetry axes fulfilling the needs of the icosahedral lattice (Morais *et al.*, 2005). For both viruses, the authors suggest that the symmetry of these hexons is due to their placement in the icosahedral web suggesting that that the needs of icosahedral symmetry overrule the skewed distortion introduced by scaffolding proteins resulting in a symmetric hexon.

In the case of the T=9 phi1026b capsid, we observed that the symmetric hexon still displays skewing. It has a 'symmetric skew' (or 'twisted skew') where three separate wedges are oriented such that one side of the wedge is 'above' the adjacent wedge and 'below' the other wedge. Further, the arrangement of the delta domains beneath the symmetric hexon is markedly different from the asymmetric hexon as the delta domains are in a singular mass as opposed to two separate towers of density. Also, the size of the symmetric hexon does not appear to suggest an intermediate state between the asymmetric hexon and the mature virion hexon. It appears that the asymmetric hexon and the symmetric hexon have two different ways for the mcp monomers to interact: the asymmetric hexon is a dimer of trimers (which provides the 'skewed' appearance) and the symmetric hexon is a trimer of dimers (which provides the 'symmetric skew' appearance). Overall, it raises a chicken and egg scenario: does the push and pull of icosahedral symmetry determine the arrangement of the mcps in the capsomer, or do the mcps and scaffolding proteins determine the

structure of the capsomers which then determine icosahedral symmetry? At least in a virus that utilizes the HK97 fold with a delta domain, the symmetric hexon still displays skewing but the skewing fulfills the icosahedral symmetry requirements of sitting on the 3-fold symmetric axis.

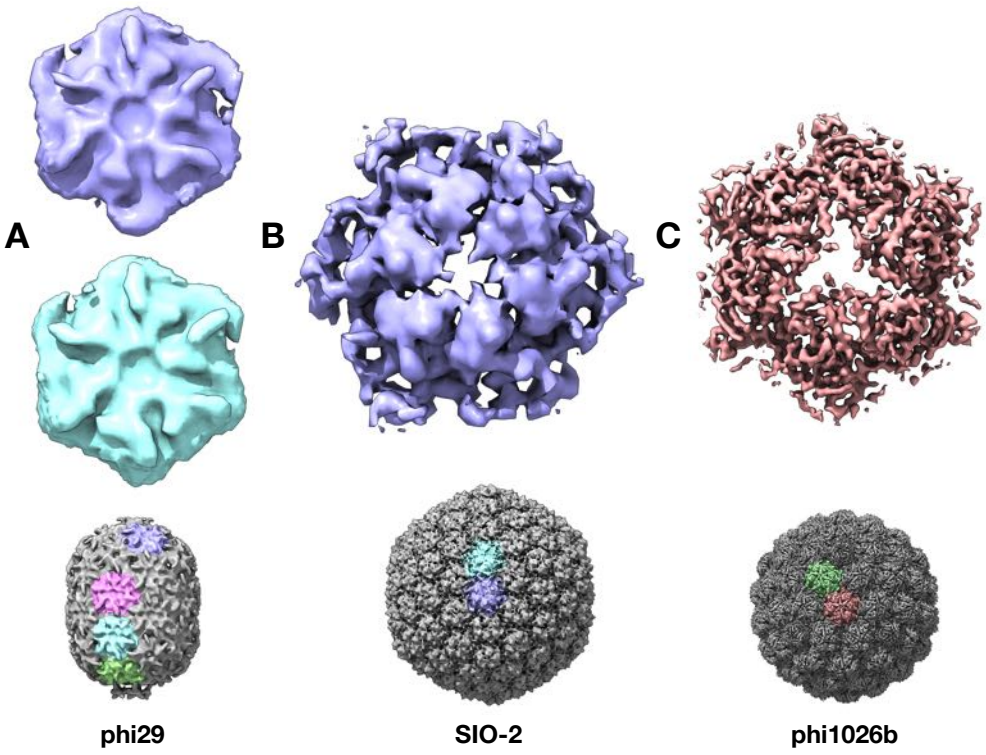

Figure 34. The symmetric capsomers in the prohead capsids of phages phi29, SIO-2, and phi1026b

A) The prohead capsid of phi29 has unique hexons highlighted in different colors for emphasis (EMDB ID: 1117) B) The prohead capsid of SIO-2 has unique hexons highlighted in different colors for emphasis (EMDB ID: 5383) C) The prohead capsid of phi1026b has unique hexons highlighted in different colors for emphasis

The angles provide an interesting insight into how phi1026b transitions between prohead and the expanded capsid.

The angles at the dimer interfaces between the phi1026b Prohead I structure and the expanded capsid undergo changes that are different than those in HK97. Table 1 shows the angles at the dimer as well as the change upon expansion. Most interesting is the unique dimer interface in the

T=9 phi1026b that is colored in magenta in Figures 24 and 25. This dimer undergoes a change of 0.9° during expansion. In HK97, only the dimer at the penton-hexon interface has such a minimal change. The magenta dimer sits at the two-fold axis. The blue dimer of HK97 sits near the two-fold axis and undergoes a large change during expansion. This shows that the two viruses manage expansion in different ways. Whereas in HK97 a majority of the change occurs at the two-fold axis, in phi1026b a majority of the change occurs near the penton-hexon interface at the green dimer. This may be a new way of gain insight into how size is managed in different viruses.

Table 1. The angles at dimer interfaces in the Prohead I and Head structures of HK97 and phi1026b

Phage	Dimer color (Fig. 24)	Chains	P-I	Head	Change during expansion
HK97	Red	G,A	71.7	74.7	3
	Green	F,B	61.3	50.9	-10.4
	Blue	E,D	45.8	68.1	22.3
	Orange	C,C	47.1	64.2	17.1
phi1026b	Red	G,A	62.7	71.2	8.5
	Green	F,B	68.6	43.4	-25.2
	Blue	E,H	56.6	52	-4.6
	Orange	C,I	66.6	55.3	-11.3
	Magenta	D,D	63.4	64.3	0.9

The difference in size of the capsomers is another way of observing how the HK97 fold is altered. The mature domain of the phi1026b mcp is 11 amino acid residues longer than the mature domain of the HK97 mcp. While it is longer, the locations that I measured for the capsomer size do not include these extra residues. 9 of the additional 11 residues in the mature domain are located in the N-arm hairpin loop. The broken interaction of E184 and R241 in the asymmetric hexon may part of a series of contacts that are altered to allow for the larger hexon.

The delta domains of phi1026b provide more insight into how the delta domains are adapted to different viruses.

In HK97, the delta domains most likely interact in two ways: as a trimer and as a penton. The delta domains of phi1026b also form a penton and a trimer but also appear to form a 6-fold coiled coil. This raises the question of whether or not the delta domains of phi1026b form the 6-fold coiled coil before or after the incorporation of the hexon into the icosahedral lattice. Either way, the delta domains have to be able to form and maintain this arrangement in addition to the trimeric and pentameric arrangements found under the other capsomers. When examining the arrangement of charges in the helical regions of the delta domains of HK97 and phi1026b, it's interesting to note that phi1026b has a far more distinct arrangement of charged and uncharged residues as compared to HK97 (Figure 35). Helical regions were selected based on predictions performed with JPred. The delta domains of the T=9 D3 also appear to have a similar arrangement of amino acid residue type to that of the phi1026b delta domains. The delta domains of the T=13 T5, which does not have a hexon on the 3-fold axis, do not have a clear grouping of charged residues which is similar to delta domains of HK97. This work is unique in not only showing skewing in the capsomer sitting on the 3-fold axis but also showing the arrangement of the delta domains beneath that capsomer.

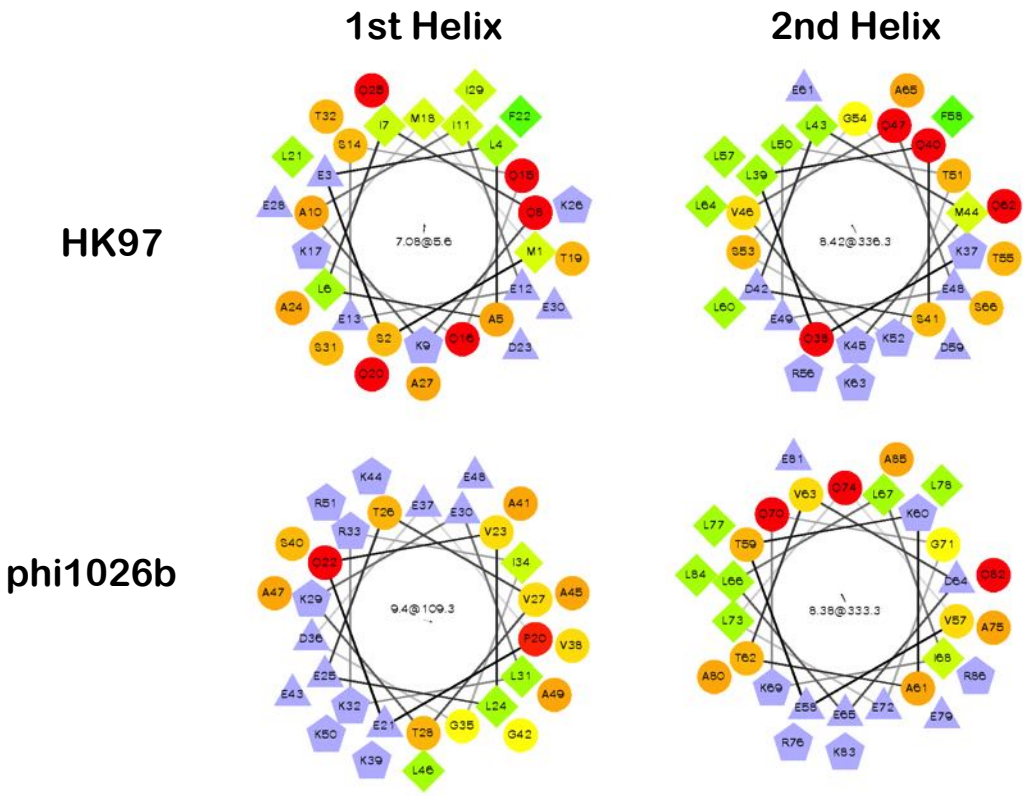

Figure 35. Helical wheels for the delta domains of HK97 and phi1026b

Helical wheels were generated using the rzlab.ucr.edu/scripts/wheel website. Shapes indicate residues types such that: diamonds are hydrophobic; circles as hydrophilic; pentagons are positively charged; triangles are negatively charged. Colors indicate degree of hydrophobicity such that: green is the most hydrophobic; yellow is zero hydrophobicity; red is highly hydrophilic; blue is potentially charged.

3.4 Future Directions

The delta domain is very hard to study due to its dynamic nature. Mutational studies may result in non-biologically relevant results as they may fundamentally change how the delta domains interact. Currently, the biophysical technique of solid state nuclear magnetic resonance (solid-state NMR) is undergoing major advances in the size of structures that can be studied (Habenstein & Loquet, 2016). For example, solid-state NMR has plumbed the structure of multiple types of fibrils (Meier *et al.*, 2017), membrane proteins with highly alpha helical assemblies (Ward

et al., 2015), viral capsids and their assembly mechanisms (Lecoq *et al.*, 2018; Suiter *et al.*, 2015), and even the structural elements of fungal cell walls (Kang *et al.*, 2018). In the last few years, the methods of cryo-EM and solid-state NMR have been combined to generate new insights into the assembly and structure of large macromolecular structures (Liu *et al.*, 2016). This new combination of technologies would be ideal to better study the highly dynamic delta domains while reducing the risk of introducing artifacts into the system by way of crosslinking or mutations. A truly ideal way to study the alpha helical bundles of the delta domain would be to apply a ss-NMR technique like a 3D $^{15}N^{13}C\alpha^{13}CX$ experiment with a long mixing time which would allow for the detection of long range contacts in the form of small dipolar couplings (Habenstein *et al.*, 2013). Further, multiple methods of cleaning up ss-NMR spectra by selective labeling such as ^{13}C spin dilution with [1-^{13}C]-glucose and [2-^{13}C]-glucose labeling have been developed over the last several years. If the prohead of phi1026b is similar to HK97 in that it can be disassembled and reassembled from capsomers, selective labeling of proheads can be even better achieved by mixing ^{13}C labeled capsomers with unlabeled capsomers and then reassembling the proheads. Combining information from ss-NMR experiments with capsid structures generated from cryo-EM can help guide atomic modeling of the delta domain.

To further understand the HK97 fold, it would be ideal to apply previously published experiments to phage phi1026b to see if the same results are obtained or if new outcomes are discovered. Similar or alternate results would provide further insights into conserved contacts that ensure assembly as well as new contacts that may allow for larger or smaller capsids. The E153-R210 salt bridge has been shown to be critical for assembly in HK97 (Hasek *et al.*, 2017) and is mostly conserved in phi1026b. When E153 and R210 are mutated in HK97, aberrant pathways resulting in tubes and sheets are generated. In some experiments, proheads are still generated but

are rare. To confirm that this salt bridge is plays a similar role in HK97, it would be worthwhile to repeat the experiment in the phi1026b system. The intra-capsomer salt bridge connection of D231 and K178 were shown to be critical to the assembly and the geometry of the HK97 prohead (Tso *et al.*, 2014). This interaction is not conserved in phi1026b and the residues are altered to smaller, hydrophobic residues (alanine and leucine) that don't appear to be in any form of contact. Would mutating these residues to those found in HK97 result in smaller capsids or aberrant pathways? While it would be challenging, repeating the experiments described in Xie and Hendrix 1995 and Li et al 2005 would also be illuminating. Finally, when HK97 Prohead I capsids are heated, it appears that the delta domains become more disordered and the hexons become more symmetrical and partially expand while the pentons and their associated delta domains remain fixed (Conway *et al.*, 2007). Such an experiment on the larger phi1026b prohead with its additional symmetric hexon may provide further insight into how the delta domains affect capsomer structure. It may also provide further insight into the stability of the phi1026b delta domains. The phi1026b delta domains are more substantial than those of HK97's, so it should take more energy to disassemble the delta domains.

4.0 The Phage λ Tail Tube

4.1 Introduction

Of the nearly 6,300 prokaryotic viruses that have been studied by electron microscopy, over 3,600 viruses are of the family *Siphoviridae* (Ackermann & Prangishvili, 2012). However, relatively little is understood about the *Siphoviridae* tail structure and assembly and much of what is understood is based primarily on studies of the bacteriophage λ tail. Assembly of the λ tail is controlled by phage genes Z, U, V, G, T, H, M, L, K, I, and J (Figure 36). The side tail fiber genes, *stf* and *tfa*, do not appear to be required for successful assembly. The tape measure protein determines the length of the λ tail (Katsura & Hendrix, 1984). The tail tip complex (TTC) proteins gpJ, gpI, gpK, gpL assemble to form a 15S complex (Katsura & Kühl, 1975), while separately, the tape measure protein, gpH, and its chaperones, gpG and gpGT, assemble (Tsui & Hendrix, 1983; Xu *et al.*, 2014). The two complexes formed by the TTC proteins and gpG, gpGT, and gpH then join with gpM to form the "initiator" (Kuhl & Katsura, 1975). Once the initiator is complete, the tail tube protein, gpV, polymerizes while displacing gpG and gpGT from the complex. The tail terminator protein, gpU, terminates gpV assembly and gpZ activates the tail in a mechanism that is currently unknown (Katsura & Kühl, 1975; Katsura & Tsugita, 1977). The tape measure protein decreases in size by ~100 amino acids between the actions of gpU and gpZ (Tsui & Hendrix, 1983). The structural changes that may occur in the tail tube protein during polymerization are poorly understood due to the lack of detailed structural studies.

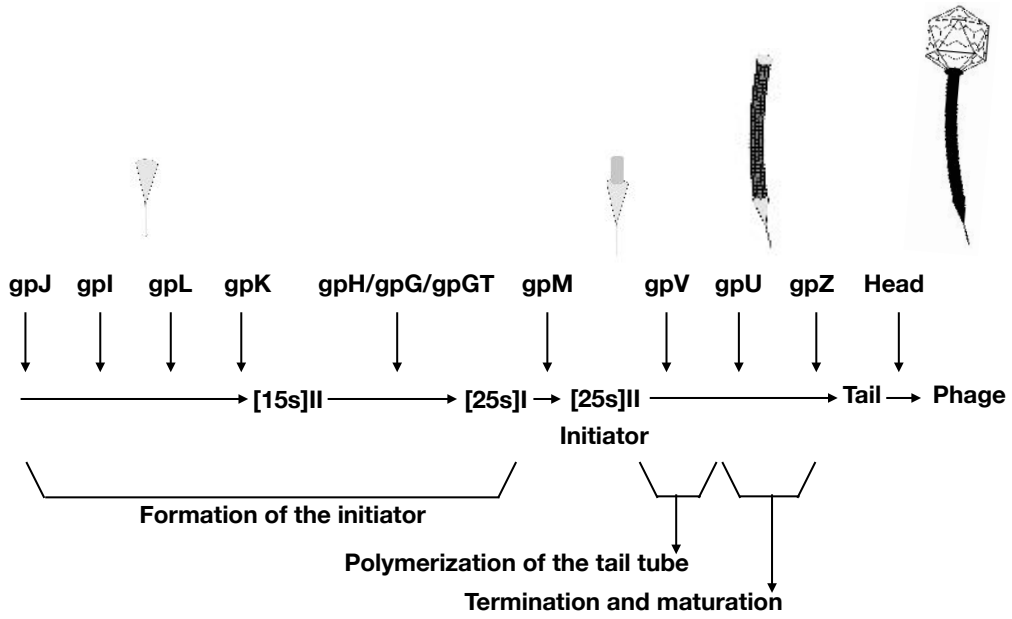

Figure 36. The assembly pathway of the λ tail

The phage genes Z, U, V, G, T, H, M, L, K, I, and J direct tail assembly in a step-wise manner. This image was adapted from Xu 2001 and Katsura 1983.

The λ tail tube protein, gpV, is one of the few *Siphoviridae* tail tube proteins that have been studied structurally. It is a 26 kDa protein with two domains: the N-terminal 'tube domain' and the C-terminal 'Ig domain' (Figure 37). When the λ tail is disassociated, individual hexameric disks appear with protruding knobs coming off a central ring (Figure 37B)(Katsura, 1981). The two domains were studied separately using solution NMR spectroscopy in their monomeric forms (Figure 37C,D). The N-terminal tube domain structure, determined with solution NMR spectroscopy (Pell *et al.*, 2009) (Figure 37C), revealed that it had structural similarity (Figure 38) with gp19, the tail tube protein of phage T4 (Taylor *et al.*, 2016; Zheng *et al.*, 2017), Hcp1 of the Pseudomonas aeruginosa Type VI Secretion System (T6SS) (Mougous *et al.*, 2006), and pyocin tube protein of a Pseudomonas aeruginosa pyocin (Zheng *et al.*, 2017). These proteins are part of tubular assemblies in other phages and bacterial systems that play a role in the translocation of both DNA and proteins (Figure 38). The conserved tail tube fold features a β-sandwich with a

large loop between β strands (the 'tail extension' loop, TE-loop). This loop is ordered in structurally similar proteins that were analyzed in the polymerized state (Mougous *et al.*, 2006; Zheng *et al.*, 2017). The presence of a helix is also conserved although it's location in the protein sequence is different between structures. Since the loop is disordered in the monomeric gpV (Pell *et al.*, 2009) and is ordered in structurally similar proteins, it has been proposed that the ordering of this loop facilitates polymerization of the proteins. The C-terminal domain of gpV features a canonical Ig-like fold of the I-set class and is a member of the bacterial Ig-like domain 2 (Big_2) family which is comprised of ~1300 phage and bacterial gene sequences (Fraser *et al.*, 2006; Pell *et al.*, 2010). The Ig-like domain is comprised of 8 β strands forming an Ig-like β sandwich with 2 β sheets comprised of 4 β strands each. Recently, a cryo-EM structure for the T5 tail tube was presented (Arnaud *et al.*, 2017) which confirmed that there is conservation of the β sandwich and helix fold for the tail tube proteins in *Siphoviridae*. Further, it showed that the long loop becomes relatively ordered in the polymerized structure. The difference between T5 and other previously studied tail tubes is that it is pseudo-hexameric since the tail tube protein is a duplication of the tail tube fold. This leaves the question of what changes occur in λ tail tube polymerization unanswered.

Another unanswered question about the *Siphoviridae* tail tube is what role, if any, do tail tubes play in genome ejection? The *Myoviridae* tails undergo considerable rearrangement during genome ejection (Leiman *et al.*, 2004; Yap *et al.*, 2016), however studies regarding the role of the *Siphoviridae* tail have not been definitive. Mutations in both gpV and gpH were able to overcome DNA injection deficiencies in *E. coli pel-* strains suggesting some role for the tail tube protein in genome injection (Scandella & Arber, 1976). The *Siphoviridae* SPP1 tail may undergo a structural rearrangement that occurs in a domino-like fashion to transmit a signal from the tail tip to the

portal for genome release (Plisson *et al.*, 2007). This rearrangement was proposed to be a result of the reorientation of two domains in the major tail protein which resulted in the reduction of the diameter of the tail tube. However, the structure of T5 tail tube proteins in tails, tails in the presence of the host receptor (FhuA), and polytails were all similar (Arnaud *et al.*, 2017). They did not observe a major rearrangement of the tail tube protein between the whole tails and the tails in the presence of the host receptor however they did not present experiments done with whole virions. They inferred that it is the opening of the tail tip and ejection of the tape measure protein that signals genome release from the head.

In this study, the polymerized form of gpV was studied by analyzing the structure of the tail tube and fitting the monomeric gpV structure into the cryo-EM density. Further, the effects, if any, of genome ejection on the tail tube protein are studied by analyzing tails purified using a plasmid expressing λ tail genes (purified tail), virion tails, and post genome ejection tails (ejected tail) to see if any observable rearrangement occurs in the λ tail as a result of DNA injection.

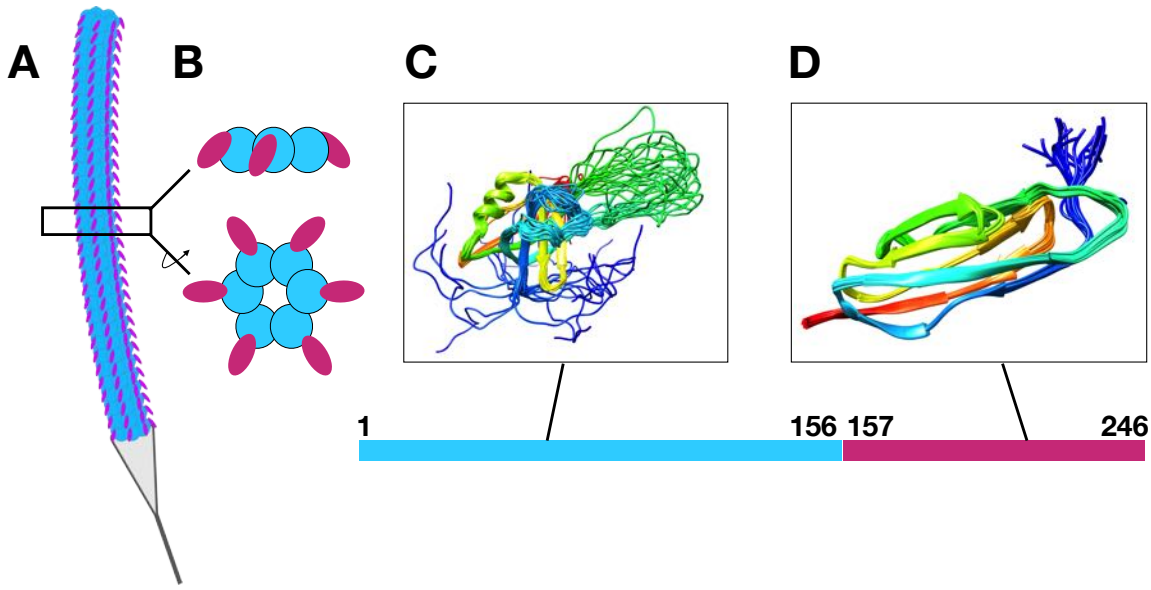

Figure 37. The λ major tail tube protein and its domains

A) A representation of the tail tube showing identical disks of the tail tube protein stacked along the length of the tail. B) A side (top) and top view (bottom) of a single hexameric disk of the tail tube protein. The two domains of the tail tube protein are colored in teal and magenta. A representation of the amino acid sequence with the same colors is to the right. C) The solution NMR structure for N-terminal domain of gpV has two unstructured regions: the N-terminus and the β2-β3 loop. (PDB ID: 2K4Q) D) The solution NMR structure for gpVC has the conserved Big_2 family Immunoglobin fold. (PDB ID: 2L04)

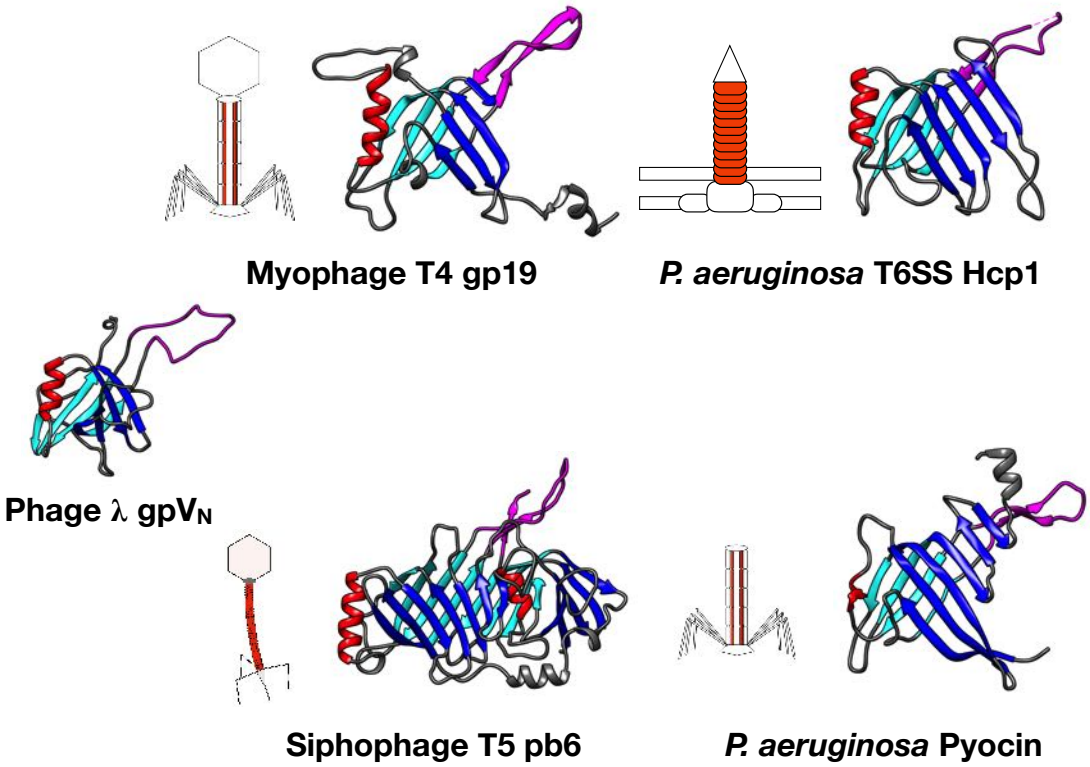

Figure 38. Proteins with structural similarity to gpV$_N$

On the far left is the solution NMR atomic model of gpVN with the distinctive features of the fold highlighted in different colors: the helix in red; the β sandwich colored in cyan and blue; the enfolding loop colored in magenta. The structurally similar proteins have the gpVN fold features colored the same. On the left of each protein is a basic representation of the structural complex they comprise. The orange red sections of the representations are where the proteins are found. PDB IDs: T4 gp19 (5W5F); P. aeruginosa T6SS Hcp1 (1Y12); T5 (*Siphoviridae*) pb6 (5NGJ); P. aeruginosa Pyocin (5W5E).

4.2 Results

4.2.1 Imaging of purified tails, virion tails and ejected tails

Biologically active λ tails were purified after expression of the λ tail genes from a plasmid as previously described (Xu *et al.*, 2014). A λ phage stock was purified using equilibrium centrifugation in cesium chloride (Bachrach & Friedmann, 1971) which yielded two widely

separated bands. The upper band contained phage "ghosts" (phage that had ejected their genomes) and proheads while the lower band contained intact mature virions.

During data collection, purified tails were imaged separately from the virions and ghost virions in order to optimize ice quality. The λ head is 55 nm in diameter whereas the tail is only 18 nm in diameter. During imaging, if the head is present, the ice is thicker due to the presence of the larger head which reduces contrast (Figure 39). Thick ice will reduce contrast, so ice should be only thick enough to fully embed the particles. Since the tails are much thinner than heads, the loss of contrast due to this effect is significant for tails attached to heads. Further, thicker ice appears to allow for more bending of the tails (unpublished observations) which reduces the number of usable images.

When imaged without the virions, the purified tail dataset has very high contrast revealing features of the tail tube such as clear striations that most likely are the hexameric tail tube disks (Figure 41). The thickness of the disks was variable due to the presence of darker spots that protrude from the tail tube on some tail tube disks which appear to be the Ig-like domains described by Katsura (Katsura, 1981) and detailed in Figure 37. Some tails have a 'hollow' appearance, whereas others appear 'full'. Most tails are the expected 170 nm in length (although some longer tails are present). Most tail tips have conical sections with striations and a protruding tail fiber that varies in appearance. At the top of the tail, one disk often appears thinner than the rest of the tail tube. We suggest that this feature may be a hexameric disk of the tail terminator protein, gpU.

Virions and ghost virions samples were mixed together before freezing specimens for imaging. The striations of the virion and ghost virion tail tubes (which will be referred to as 'ejected tails') were not as well-defined as those in the purified tail sample (Figure 41). The virions had the appearance of 'full' tails and ghost virions had the appearance of 'hollow' tails. The hollow ghost

virion tail tubes (ejected tails) were similar in appearance to those observed in the purified tail sample. In both cases the protrusions from the tail tube were still apparent. The ghost virions did not appear to contain DNA and many did not have well-ordered tail tips, as might be expected for virions that had ejected their genomes.

The power spectra from the purified tail, virion tail and ejected tail datasets had similar helical parameters (Figure 40B). All three samples have a meridional at the same approximate height suggesting the same helical rise parameter. The first layer line also occurs at approximately height suggesting approximately the same rotation for all three datasets as well. In all three samples, the layer lines of the first Bessel order have merged on the meridian suggesting out of plane tilting (Desfosses *et al.*, 2014).

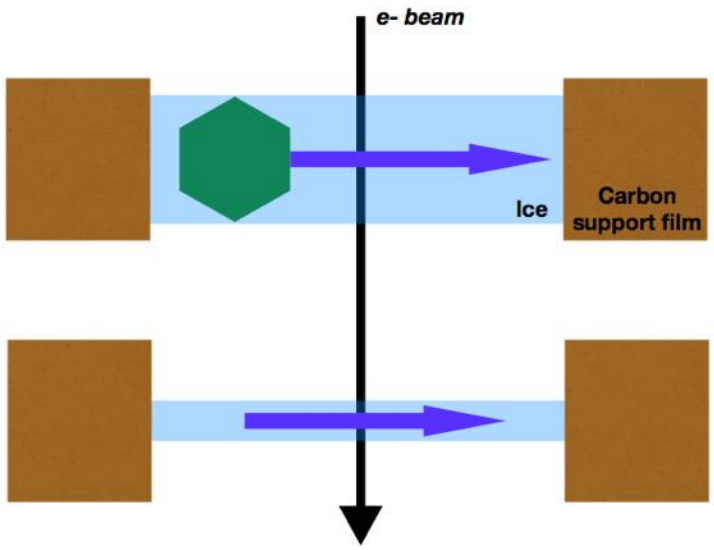

Figure 39. Imaging purified tails separately allows for higher contrast

The packaged head of the virion has a much greater diameter than that of the tail and causes thicker ice.

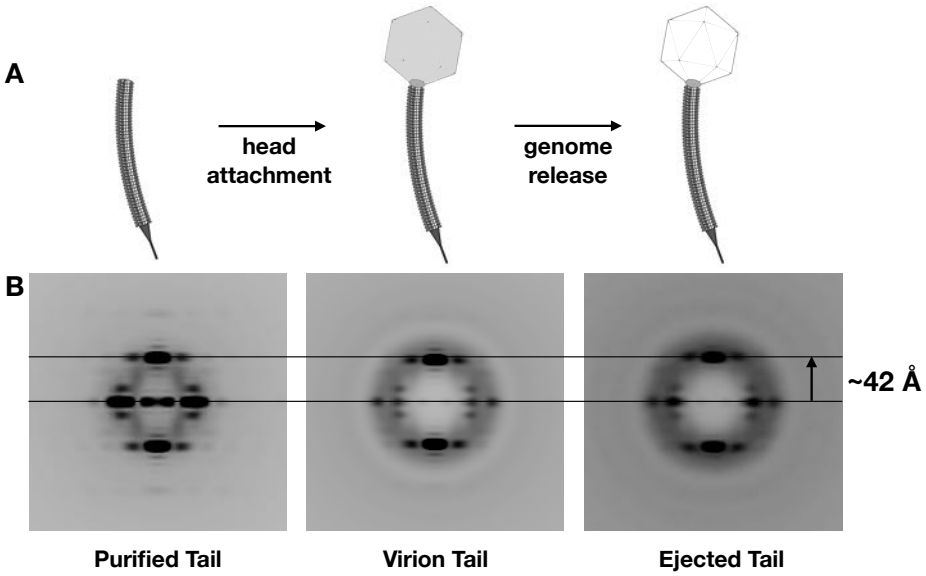

Figure 40. The power spectra of the purified tail, virion tail, and ejected λ tail datasets

A) A basic schematic showing the relationship between the three samples. B) Average power spectra generated from the three datasets show similar helical parameters like rise and rotation.

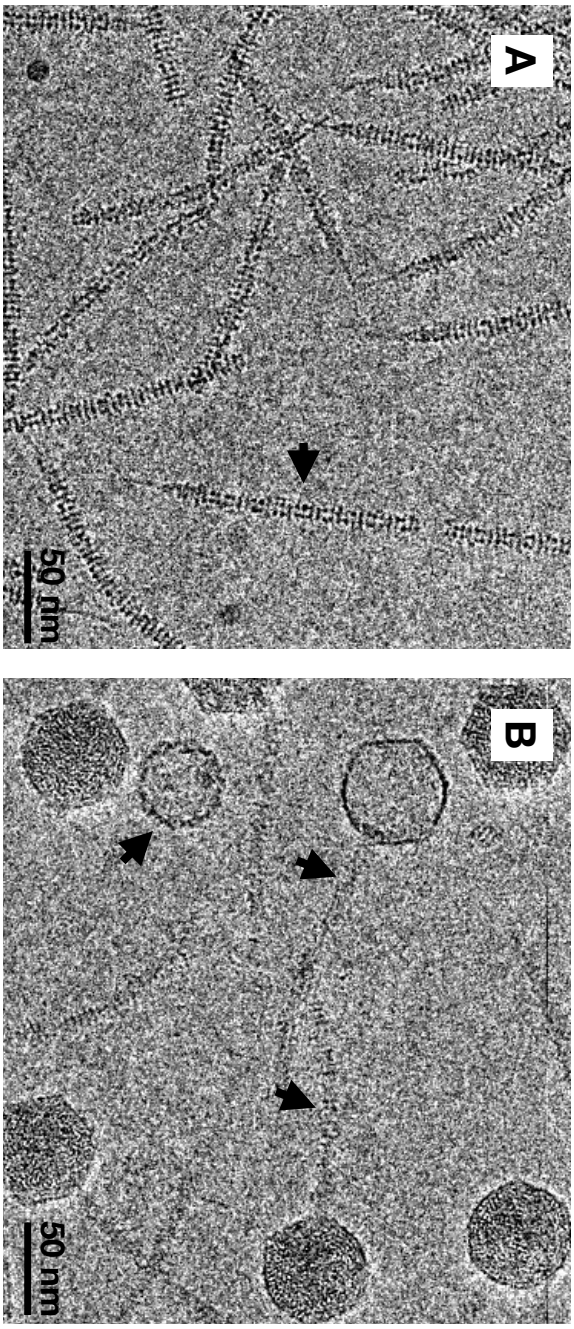

Figure 41. Cryo-EM micrographs of the purified tails, virion tails and ejected tails

A) A micrograph showing the purified tails. Arrows indicate clear striations of the tails. B) A micrograph showing the combined sample of the virions and ghost virions. Arrows indicate tails with visible striations as well as a prohead.

4.2.2 Tail tube structures of purified, virion, and ejected tail tubes

The purified tail tube reconstruction resulted in one of the first sub-nanometer structures of a *Siphoviridae* tail tube with a resolution of ~7 Å (Figure 42) with over 25,000 particles. The resolution of the ejected tail tube has the lowest resolution which is around ~11 Å. This dataset was relatively small compared to that of the other two datasets. The virion had a resolution of ~9 Å with nearly 32,000 particles. The separate software packages Spring, RELION, and IHRSR++ were used. Results from Spring and Relion yielded higher apparent resolution than those of IHRSR++, likely due to the limited size of the datasets that the version of IHRSR++ could accept. 6-fold symmetry was applied based on observations of dissociated tail tube disks (Katsura, 1981). Multiple helical parameters were also tested that ranged from 42.4 Å to 44 Å for rise and from 17.5° to 19° for rotation.

Only the virion had internal density which we attribute to the tmp (Figure 43, column 3). The tmp density appears partially segmented with boundaries that match the disk separation and is likely to be an artifact of the helical reconstruction process. The internal density in the virion does not appear to contact the walls of the tube. While we expected to see internal density in the purified tail and ejected tail tubes, both appeared hollow. In both cases, the density within the tube appears to be background noise that has been amplified by the reconstruction software. The interior walls of all three structures are contiguous, curved and featureless with the exception of the gap between disks. Column 4 of Figure 43 is a top view of the reconstructions showing no major changes in tube diameter between the three samples.

The structures have clear, defined disks along the length of the tube. The disks of the three reconstruction appear as thick bands along the tube matching what is seen in the cryo-EM images. These disks have several distinct features that are only evident when the threshold for surface

rendering was varied from high to low values. At higher thresholds, large semi-pyramidal structures protrude from the tube and are shown in Figure 43, column 5 highlighted with magenta. These match the protrusions radiating from dissociated tail tube disks, that were attributed to the C-terminal Ig-like domain of the tail protein (Katsura, 1981; Pell *et al.*, 2010). In the purified tail and the virion samples the Ig-like domains appear to contact the Ig-like domains on the disks above and below. In later reconstructions, the density corresponding to the relatively flexible Ig-like domains were cut off by reducing the diameter of the reconstruction in order to focus on the more rigid tube domain and achieve higher resolution.

The strongest density in the structure forms a tube that is 9 Å in diameter which corresponds to previous observations of the λ tail tube disk (Katsura, 1981). In column 2 of Figure 43, the density corresponding to the tube domain is shown in teal. The striations that were visible in the cryo-EM micrographs are shown to be the individual hexameric tail tube disks. Overall, each disk has a wide base with a narrower top with round knobs at high radius that are in the same location as the Ig-like domains which appear to be a part of the tube domain due to the strength of the density. Next to the knobs in the purified tail and virion structures are longitudinally arranged rods of density that appear to correspond to the helix found in the gpV_N NMR model. All three tail samples show only weak density between disks (Figure 43) with wisps of strand-like density weakly connecting disks together. This inter-disk density, while weak and somewhat variable in the maps, was strongest in the purified tail reconstruction.

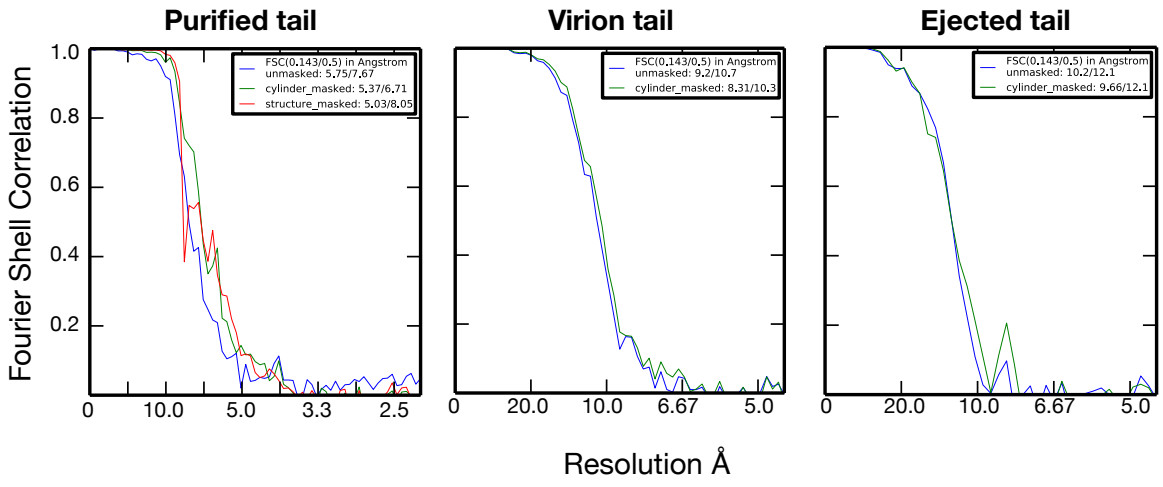

Figure 42. FSC curves for the plasmid tails, virion tails, and eject tails from the Spring software suite

4.2.3 Visualizing the gpV fold in the density maps

The features of the gpV fold are most apparent in the purified tail reconstruction which reached a resolution of ~7 Å. Comparisons to T4's gp19 make it clear that the helix forms the external face of the tube. Further, there is a distinct rod-shaped density that will fit the helix confirming that the helix is on the outer face. In addition to the rod-shaped density is a pronounced 'knob'-shaped density that can fit the β1-β2 loop ('outside' loop, O-loop) (Figure 44).These two features help orient the 'inside and outside' features of the tube, however, the orientation of the atomic model in the density is not immediately apparent. A single monomer of the solution NMR gpV atomic model (PDB ID: 2K4Q) can be oriented such that the helix is either to the left or the right of the O-loop knob (Figure 46). The connection from the helix to the 4th β sheet fits into well-defined density in both orientations which are highlighted by the insets in Figure 46. The more ideal orientation becomes clear when viewing the interior of the tube. The 6th and 7th β sheets fit

the density better when the helix is to the left of the O-loop. The inset on the bottom left in Figure 46 shows how the β sheets fit into two distinct outcrops from the wall. In the alternate orientation these outcrops are not occupied by the model.

The distinctive 'β-sandwich and helix' of the gpV atomic model fits into the virion tail tube density. Figure 45 shows the solution NMR gpV structure rigid body fitted into the purified tail and virion tail tube densities. The distinctive features of the helix and O-loop are also visible in the virion density. These allow for the orientation and rigid fitting of the gpV monomer into the density. In both structures, once the structure is fit into the density, the N-terminus, C-terminus and β2-β3 loop ('tail extension' loop, TE-loop) is not occupying density. Further, while the β-sheets of the β-sandwich fit the density rather well, some adjustment may be necessary.

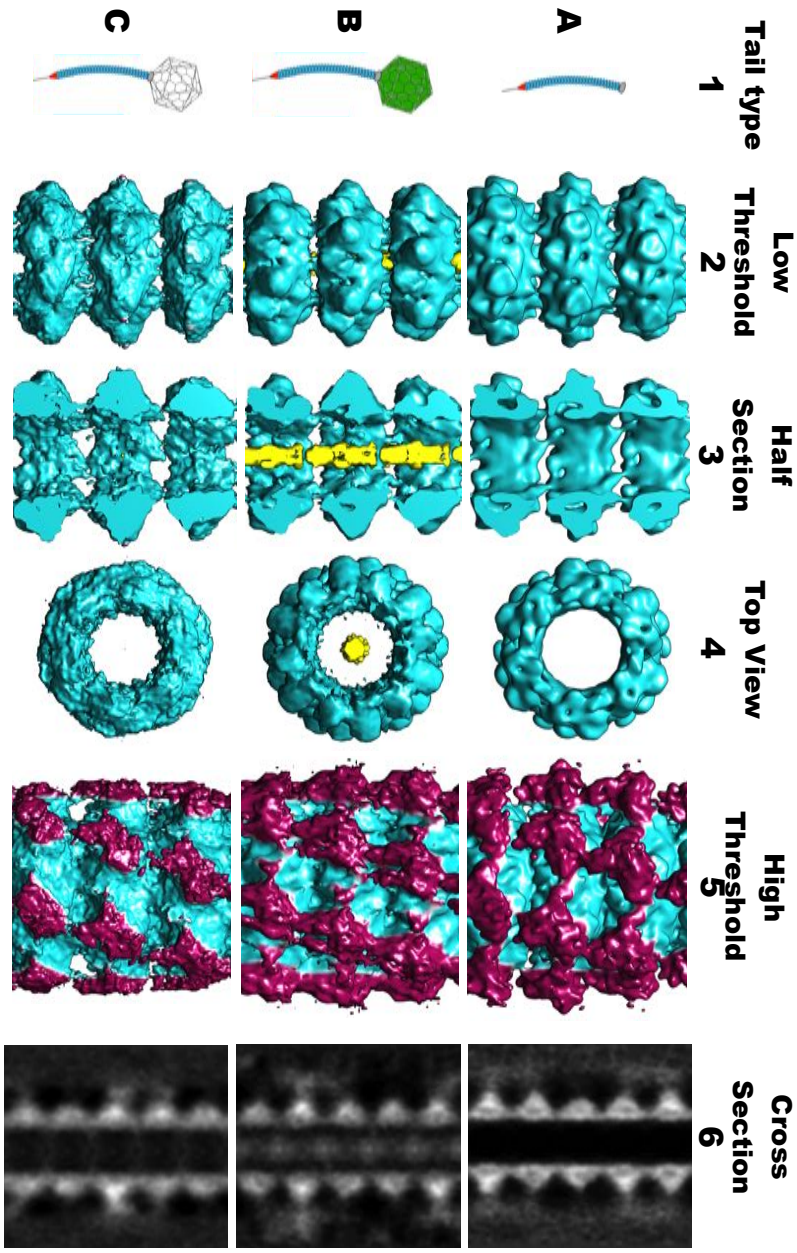

Figure 43. Reconstructions of plasmid tails, virion tails, and eject tails

Column labels detail how the reconstruction is being presented. Rows are based on tail tube type. Colors correspond to either the domains of gpV (teal for the tube domain and magenta for the Ig-like domain) or internal density (yellow). The rows are A) free tail results, B) virion results, and C) ghost tail results.

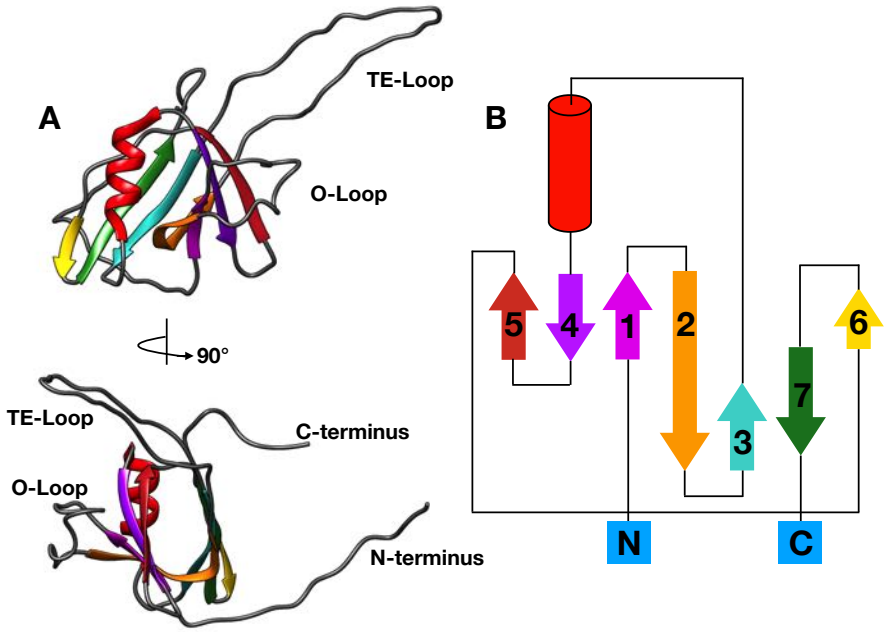

Figure 44. The λ gpV tube domain

A) Secondary structural elements are colored and labeled. B) A 2D schematic with corresponding colors for the secondary structural elements (PDBsum).

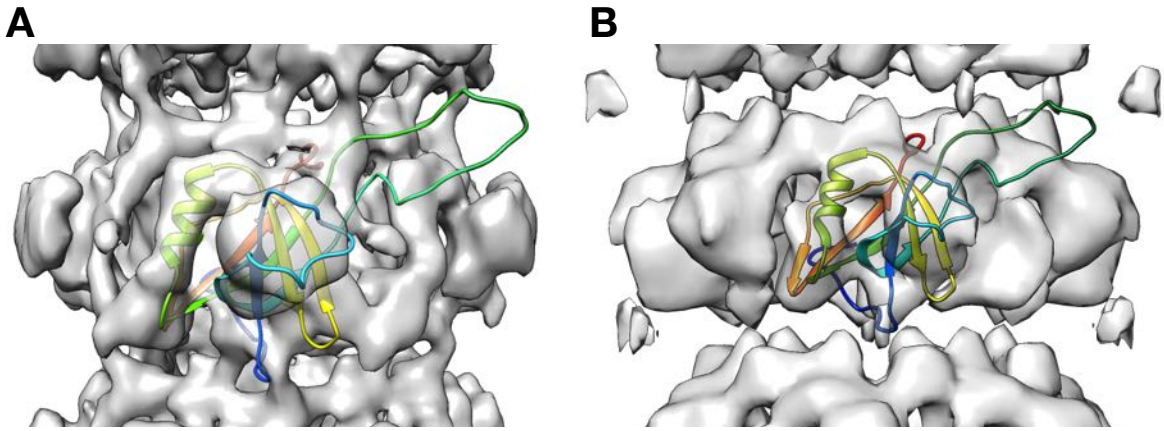

Figure 45. Rigid fit of the gpV NMR structure into the plasmid tail (A) and virion tail (B) density maps

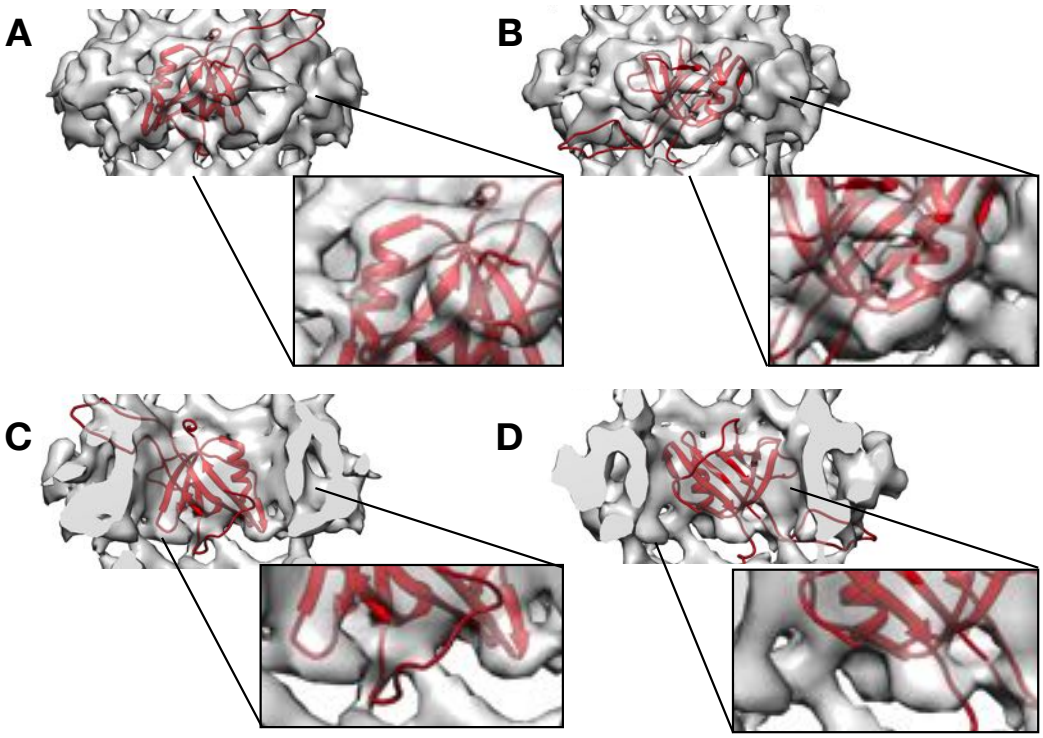

Figure 46. Alternate fitting of the gpV$_N$ atomic structure

A,C) The external and internal view of the placement of the gpV atomic model with the helix to the left of the O-loop. B,D) The external and internal view of the placement of the gpV atomic model with the helix to the right of the O-loop.

4.2.4 Adjusting the monomeric gpV structure to fit the cryo-EM density of polymerized gpV

The orientations of the termini and TE-loop of the solution NMR structure of gpV result in considerable steric hindrance with adjoining monomers. Figure 47 illustrates the different orientations of the termini and TE-loop in the polymerized T4 tail tube and rigid body fitted gpV monomers in the tube density. The N-terminus, C-terminus, and TE- loop are colored in red, green and blue respectively. As stated earlier, rigid body fitting of the gpV NMR model resulted in these three parts of the structure not occupying density. The termini were located inside the tube whereas

the TE-loop was located outside the tube density. Since the C-terminus of the tube domain connects to the external Ig-domain it cannot be located inside the tube. Figure 47 shows how the termini and the TE-loop help coordinate the connections between tube disks in the T4 tail tube. The orientations of the termini of the TE-loop in the gpV solution structures does not allow for connections between the tail tube disks (Figure 47). Further, the current orientation of the gpV TE-loop does not allow room for an adjoining monomer. Altogether, this suggests that the orientations of the termini and TE-loop are not conducive to a polymerized form of gpV and need to be adjusted.

The T4 tail tube (gp19) model suggests how the termini and TE-loop of gpV_N need to be adjusted (Figure 47C,D). The super position of the two structures showed that the termini of gp19 are located on the exterior surface of the tail tube whereas the termini of gpV are not. The enfolding loop of T4 is oriented towards the interior of the tube as well as moving away from the β-sandwich resulting in an interface where an adjoining monomer can fit. The green C-terminus needs to be moved to the external surface of the **λ** tail tube. This will allow for the connection of the C-terminus of the N-terminal tube domain to the N-terminus of the C-terminal Ig-like domain which is on the external surface of the tail tube (Katsura, 1981). The red N-terminus will need to be adjusted so it is located on the external surface of the tail tube as well. The TE-loop in the gpV_N atomic model will have to be adjusted to the interior of the tail tube allowing room for an adjoining tail tube monomer (Figure 47C,D).

Flexible fitting of the gpV atomic model was performed using interactive molecular dynamics flexible fitting (IMD-MDFF), molecular dynamics flexible fitting (MDFF), and Phenix. Since the termini and the TE-loops were so far away from where they needed to be, it was imperative to use the tool of IMD-MDFF to guide the model to where they fit best. Not using this

tool would have made a correct fit highly unlikely. Once the termini and the TE-loop were oriented correctly, a regular MDFF trajectory was performed for additional fitting of the entire fold into the density. Phenix was used to further investigate how the fold could fit into the density. As the resolution of the density was not high enough to gauge connectivity of the structures, I utilized more than one fitting tool to see how the structures might occupy the density. The final atomic model was generated by initial guided fitting in IMD-MDFF followed by an additional undirected MDFF run and further refined using Phenix.

Flexible fitting the termini and TE-loop generates a more compact structure similar to the tail tube disk. Figure 48 shows both a side and top view of the solution structure of gpV rigid fitted into the density and gpV flexibly fitted into the density. The top view shows how moving the termini to the exterior of the tube removed structures from the interior of the tube. This matched the reconstruction results for the purified tail tube and ejected tail tube since no internal density in those reconstructions was observed. This also matched the virion tail tube since there were no resolved connections from the tube density to the tmp interior density. The O-loop also fits the density better as the loop was extending past the density when rigid fitted allowing for a more compact structure overall. As shown in Figure 48, moving the TE-loop allows room for an adjoining monomer. The TE-loop 'reaches over' the adjoining monomer. The loop extends far enough that it is possible that the monomer makes contact with the N+2 monomer which would allow for better connectivity throughout the hexamer.

While the C-terminus and TE-loop fit the density, density that would accommodate the N-terminus was not visible. Unlike the rest of the gpV monomer, the first 10-12 residues are not well defined in the density. Once it became clear that placing the first 10 residues would become subjective, I removed them from the model. What remained of the N-terminus was oriented in a

position similar to that of the T4 gp19 fold and the truncated N-terminus reaches into the density that would be mostly occupied by the adjoining monomer's helix (Figure 48D). This appears to be the best location for the truncated N-terminus as no other density is unoccupied.

The β-sandwich and helix also appear to undergo a degree of change to accommodate the density map. The β sheet comprised of β1/β2/β4/β5 appears to reorient relative to the β sheet comprised of β2/β3/β6/β7. The orientations of the β-sandwich β sheets in the flexibly fitted gpV_N match those of T4's gp19 (Figure 49) suggesting that this reorientation is meaningful. This reorientation of β sheets was not immediately apparent during rigid fitting of the solution NMR structure into the density map. The β turns generated by β4-β5 and β6-β7 undergo the most change in the β-sandwich along with 2nd β strand by moving over 12 Å. This is obvious when comparing the RMSD of the α carbons of the unfitted and fitted atomic models for gpV (Figure 50). Both of these β turns sit at the interface between adjoining monomers suggesting that these need to rearrange to allow for the hexamerization of gpV. The helix also rearranges during polymerization as well. Once gpV_N is flexibly fitted the orientation of the helix matches that of T4's gp19 (Figure 49). Along with the rearrangement of the TE-Loop, this suggests a conserved mechanism of polymerization among structurally similar proteins.

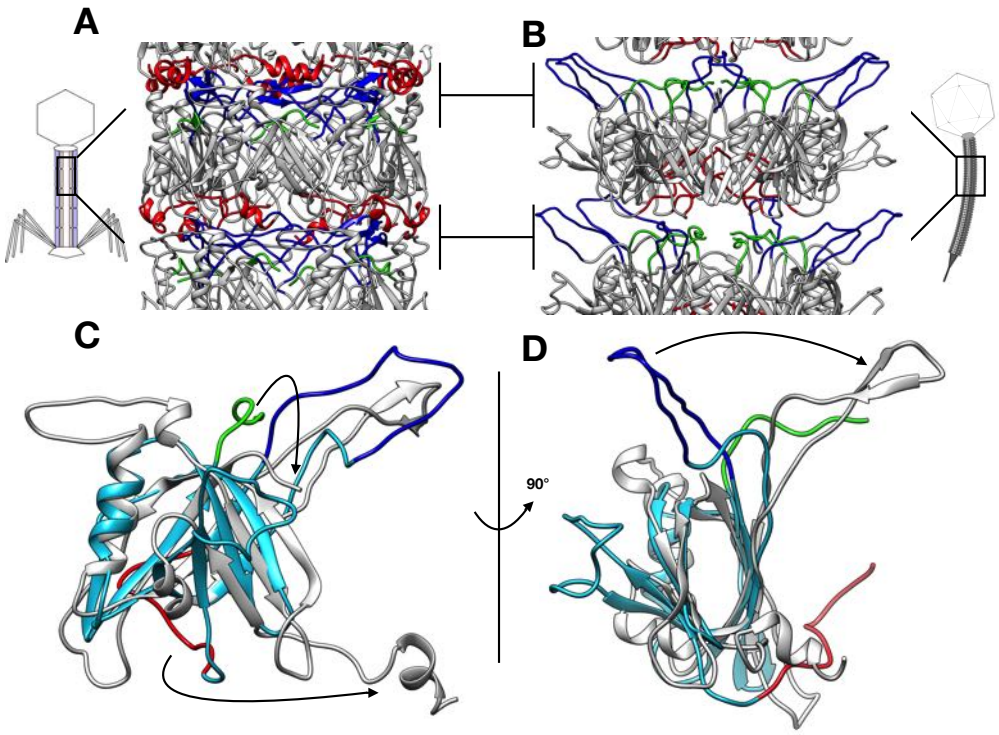

Figure 47. The TE-loop changes during polymerization based on the homologous T4 tube protein gp19

A) The atomic model of the T4 tube protein gp19 showed how the loops (blue) and N- and C- termini (red and green respectively) organize during polymerization. (PDB ID: 5IV5) B) The solution NMR arranged into hexamers show the TE-loops are facing outward and are not in contact with neighboring hexamers. C-D) The atomic models for T4 gp19 (grey) and gpV (teal with highlighted regions) are superimposed. The arrows indicate how different regions might change during polymerization.

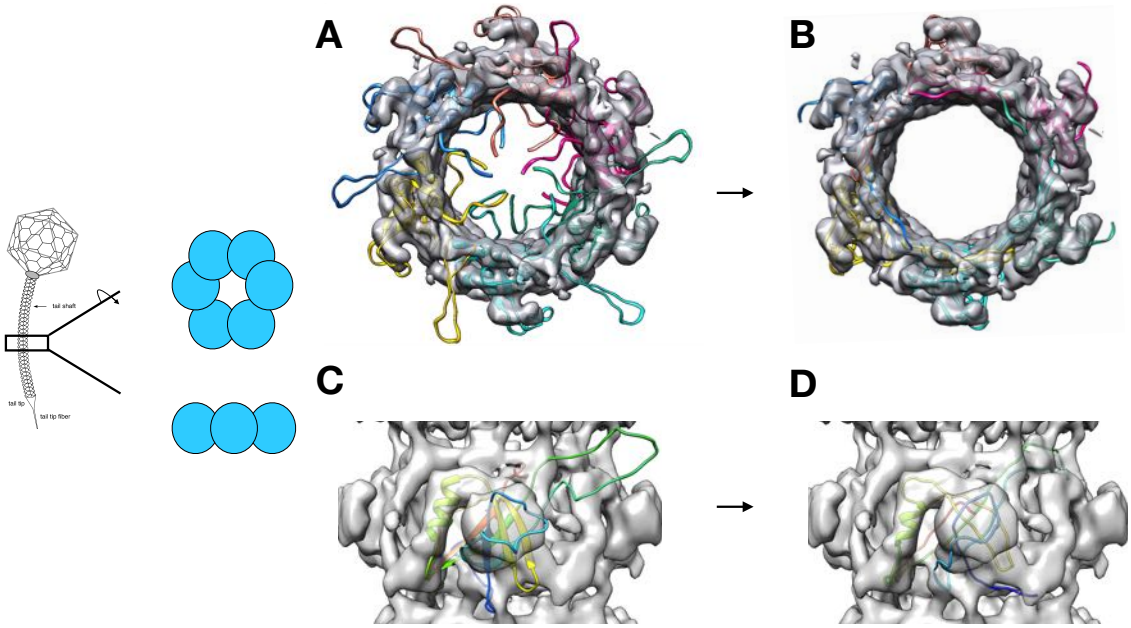

Figure 48. Rigid and flexible fitting of gpV$_N$ into the cryo-EM density

In all the structures the density corresponding to the Ig-like domain has been removed. The first 10 residues of the N-terminus are removed during flexible fitting. A) A top view of a single hexamer with the gpV$_N$ structure rigid fitted into the cryo-EM density. B) A top view of a single hexamer after the gpV$_N$ structure has been flexibly fitted into the density. C) A side view of a single hexamer with the gpV$_N$ structure rigid fitted into the density. D) The same view as C but after flexible fitting.

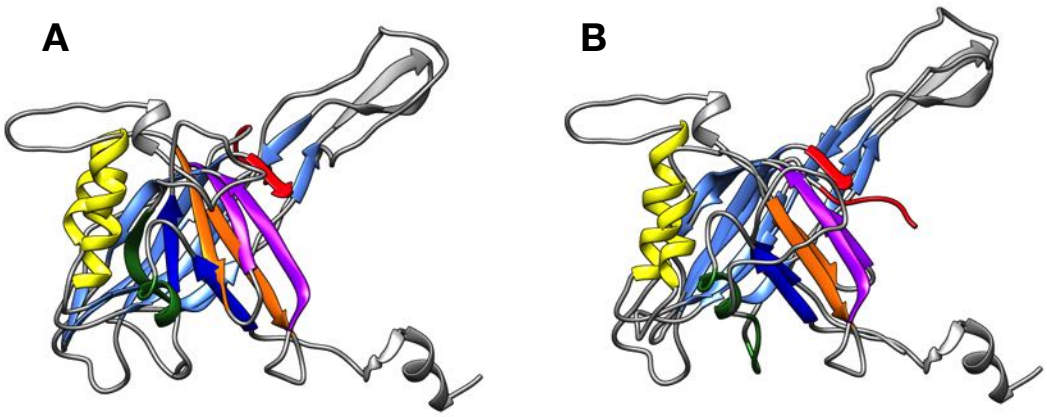

Figure 49. Rearrangement of the β sheets and helix during fitting matches orientations of the T4 β sheets

A) The solution NMR structure of gpVN is superimposed on the structure of T4's gp19 with the helix and β strands highlighted for emphasis. B) The flexibly fitted structure of gpVN superimposed on T4's gp19 with β strands and helix highlighted for emphasis.

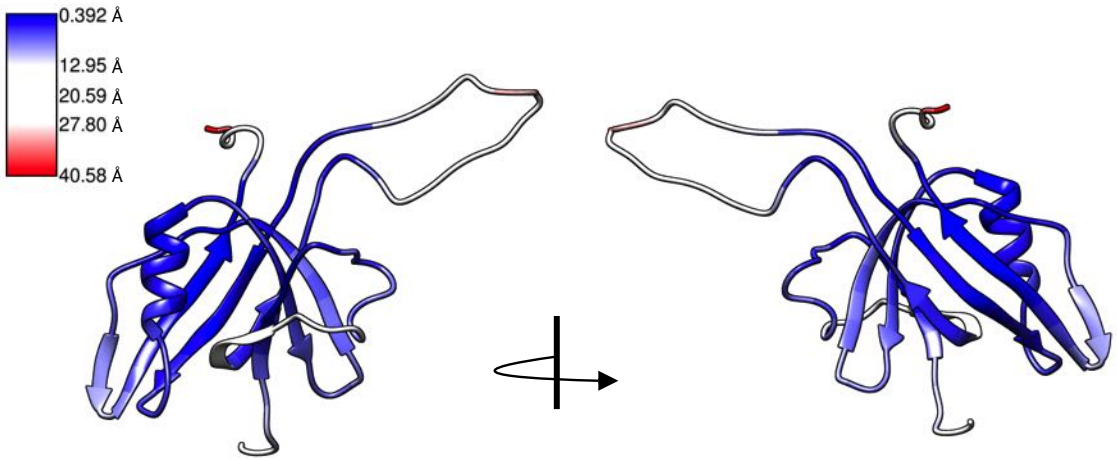

Figure 50. Degree of changes that occur in monomeric gpV$_N$ during flexible fitting

The monomeric form of gpVN is colored according to the degree of change as measured by the RMSD of the α carbons.

4.2.5 The tape measure protein appears to be unstable in the plasmid tail sample during storage

The tail tubes of the purified tail sample appeared either 'hollow' or 'full'. Purified tails were imaged with regular cryo-EM and cryo-negative staining. In cryo-EM images, there were hollow and full tails as well as rare tails with unusually long tail fibers (Figure 52). In tubes that appeared hollow, a clear channel connected from the top of the TTC to the end of the tail tube. The tail fibers on the tail tips were short with a 'ball' at the end which match observations in the literature (Hohn & Katsura, 1977; Katsura, 1983), or were unusually long. In cryo-negative stain images, many of the tails had a hollow appearance with a clear channel running from the top of the TTC complex to the end of the tail tube (Figure 51). In some cases, the channel extended through the TTC. Rarely, this channel was not present and the tails had a full appearance (Figure 51). The presence of stain inside the tail tube implies that there was no structure impeding the flow

of the stain into the tube suggesting that the hollow appearance is not merely a visual artifact. While gpH is expected to fill the entirety of the tail in an extended state, these images suggest that gpH was not stable in the purified tail sample tail tube.

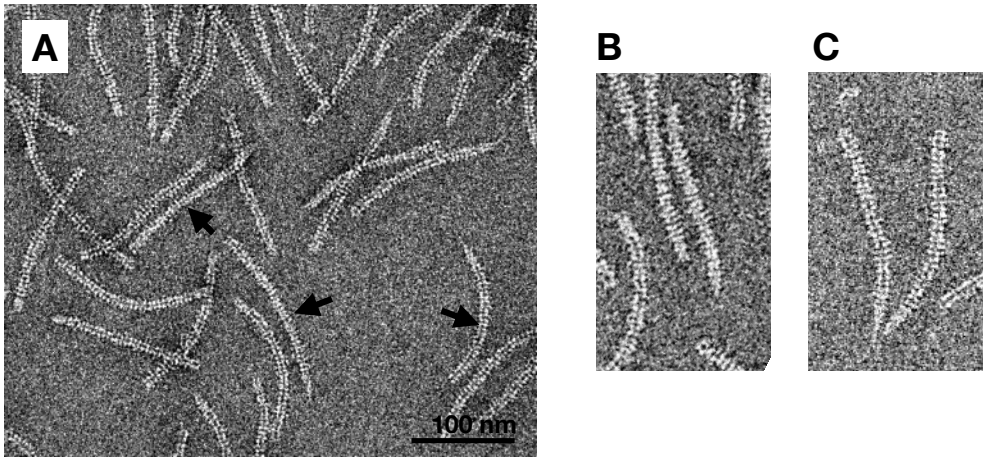

Figure 51. Cryo-negative staining of plasmid tails show an internal channel

A) A micrograph showing that most of the tails have a visible internal channel. Arrows mark tails that do not appear to have a channel. Representative images of tails without an internal channel (B) and very clear channels (C) are shown.

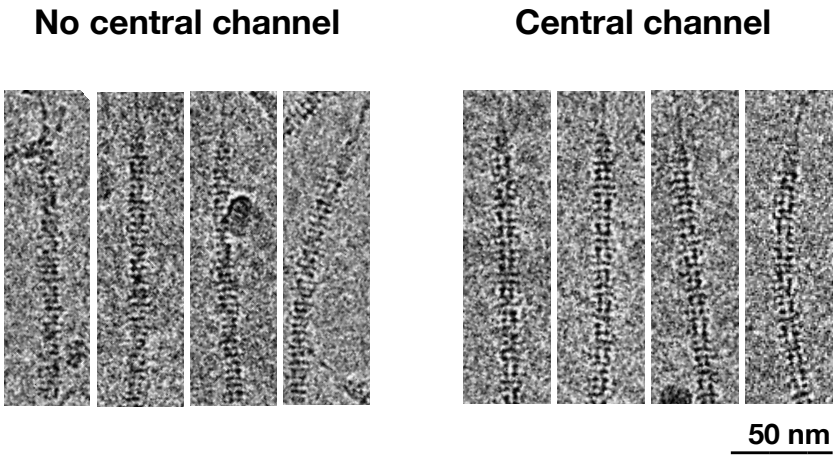

Figure 52. Cryo-EM images of plasmid tails show an internal channel

On the left, purified tails that have no visible internal channel. In the center, tail that have a partial or complete internal channel and unusually long tail tip fibers. On the right, tails that have a visible internal channel and normal looking tail tips.

4.2.6 Variability in the orientation of the gpV Ig-like domain

The electrostatic surface potential of the Ig-like domain suggests one orientation of the Ig-like domain relative to the tube domain. Figure 53 shows the external surface of the tube domain as well as the surfaces of the Ig-like domain. The face of the Ig-like domain that is comprised of β4/β5/β6/β7 has a distinct patch that is positively charged. The N-terminus of the Ig-like domain is also oriented towards this patch. The opposite face is which is comprised of β1/β2/β3/β8 is more negatively charged than the other side. The O-loop and C-terminus of the tube domain are negatively charged. Based on the electrostatic surface potentials the positively charged surface of the Ig-like domain is in some form of contact with the negatively charged area of the tube domain. This potentially explains why the Ig-like domain and the O-loop knob of the tube domain are in the same location.

The density allows for multiple orientations of the Ig-like domain (PDB ID: 2L04) when it is connected to the adjusted model of the gpV tube domain. In the program Chimera, I connected the C-terminus of the adjusted model of the gpV tube domain to the N-terminus of the Ig-like domain structure. Once connected, I adjusted the Ig-like domain to fit the Ig-like domain density. Doing this I found multiple orientations the Ig-like domain could adopt. Figure 54 shows how a connected Ig-like domain can remain in register with its corresponding monomer (orange and blue), reach over to the adjoining monomer (sea green), or 'reach' up to the above disk (magenta). All four orientations fit the density relatively well and allow for the alignment of the positive charged face of the Ig-like domain with the negatively sharged O-loop knob of the tube domain. Despite this potential variability in orientation, the Ig-like domains may make some form of

contact with the Ig-like domains of disks above and below (Figure 54/Figure 43 column 5). This appears to be most clear for the purified tails and the virions.

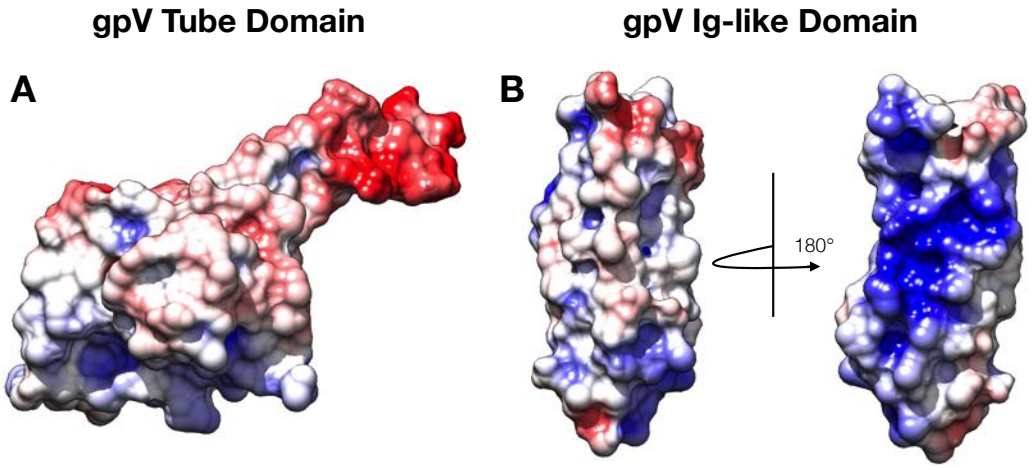

Figure 53. Electrostatic surface potentials of the tube domain and Ig-like domain

A) The electrostatic surface potential of the outward facing surface of the tail tube domain is shown. B) The two sides of the Ig-like domain electrostatic surface potential is shown. The color gradient from red to blue represents negative (−5 kT/e) to positive electrostatic potentials (+5 kT/e). Surfaces calculated using APBS from NCBS.

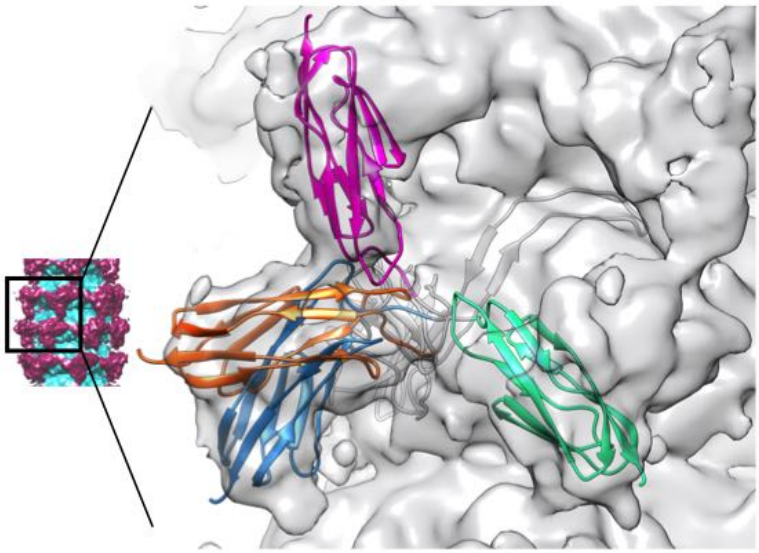

Figure 54. A model for how the Ig-like domain may fit into the tail density

Four copies of the Ig-like domain atomic model are connected to the tube domain (grey) illustrating how it's possible for the Ig-like domain to fit into the density in different ways.

4.3 Discussion

Phage tails that are structurally similar to the λ tail tube provide insight into λ tail tube flexibility

Several differences between the λ tail tube and the tail tubes of T5 and T4 might explain how and why the λ tail tube is so flexible and why higher resolution structures are not possible at this time. First, the λ tail tube disks have less overlap than structurally similar tail tubes of T4 and T5. A striking feature of the λ tail tube reconstructions is the relative lack of density connecting tail tube disks. In T5, there is more overlap between the disks compared to λ. Figure 56 shows how the T5 enfolding loops come in contact with a helix associated with the above disk. The cryo-EM density for the λ tail tube does not suggest any type of overlap as seen in T5. In Figure 56, T5 has a well-defined helix in its N-terminus that sits at the junction between disks. When the tube proteins of T5 and gp19 of T4 are aligned, this helix is almost aligned with the N-terminal helix of gp19. If λ had a more stable N-terminus, it might also be sitting in this location.

The tail tube of T4 has even better connectivity between tail tube disks and is far more rigid than λ (Figure 55). The T4 gp19 tail tube is more of a continuous density without clear delineation of the separate disks unlike the clear striations seen in micrographs of λ tail tube. Figure 55 shows how at the junction between two disks of gp19 at least 6 monomers are in contact. The λ tail tube density suggests no such interconnectivity that is very stable. The N-terminus of gp19 is a well-defined helix which is not seen in λ. Also, there is an additional loop between the helix and the 4^{th} β sheet in gp19 that is not present in both T5 and λ. This loop is nearly 15 residues long and makes contact with the adjoining monomer. Further, it connects to an enfolding loop of a monomer that is 2 positions away as well as an N-terminus of another disk that is also 2 positions away. Similar connections are not made in the λ tail tube. Further, inside the tail tube, the adjoining

monomers of T4 gp19 make a continuous wall comprised of β-sheets (Figure 55D). The density suggested that within the λ tail tube, adjoining monomers have nearly 17 Å of continuous contact whereas T4 has nearly 26 Å of continuous contact. Also, there is an additional loop between B6-B7 that is not present in λ that is nearly 6 residues long. This allows for a connection to the N-terminus of an adjoining monomer as well as the helix loop and enfolding loop of two separate monomers in the next disk. Once again, this connection is not present in λ.

The lack of definition of the N-terminus may explain tube flexibility. The first 10 residues were removed from the gpV model due to a lack of connectivity in the density. This was surprising considering that the N-terminus of T4 plays a role in tube interconnectivity (Figure 55C). One explanation for the lack of definition is that the N-terminus features 4 prolines in a 12 residue stretch. In the NMR structure, this stretch of residues had no long range 1H-15N heteronuclear NOEs (Pell *et al.*, 2009). It appears that even in the polymerized form on the tube, it may remain unstructured. They also do not feature a strong electrostatic charge (it is generally neutral) compared to the top and bottom of the tail tube disks which feature stronger net charge. Some lower resolution reconstructions showed an extra piece of density that seemed to contact the Ig-like domain.

Finally, the helical parameters of λ are different than those of structurally similar proteins which may show the effect of different disk connectivity. While the helical reconstructions for all three datasets converged at different helical parameters, the helical rise varied from 42.1 Å to 43.5 Å and the rotation varied between 17 and 19 degrees. Since the tail tube is flexible, it is possible that populations of particles with alternate helical parameters exist. Datasets containing populations that differ by as little as 1 degree and 1 Å have been reported on before (Egelman, 2007) further emphasizing the flexibility of the λ tail tube. The helical rises of T5 (*Siphoviridae*)

and the phage T4 (*Myoviridae*) are 40.6 Å and 40 Å respectively. The same is true for phage SPP1 (*Siphoviridae*) (Langlois *et al.*, 2015; Zinke *et al.*, 2017) which has a rise of 40 Å. The rotations of these structurally similar proteins were similar to that of λ with rotation of ~18 degrees for both. The reason for this difference is not immediately clear. Both T5 and T4 do have extra loops and extensions that are not present in the λ tail tube domain.

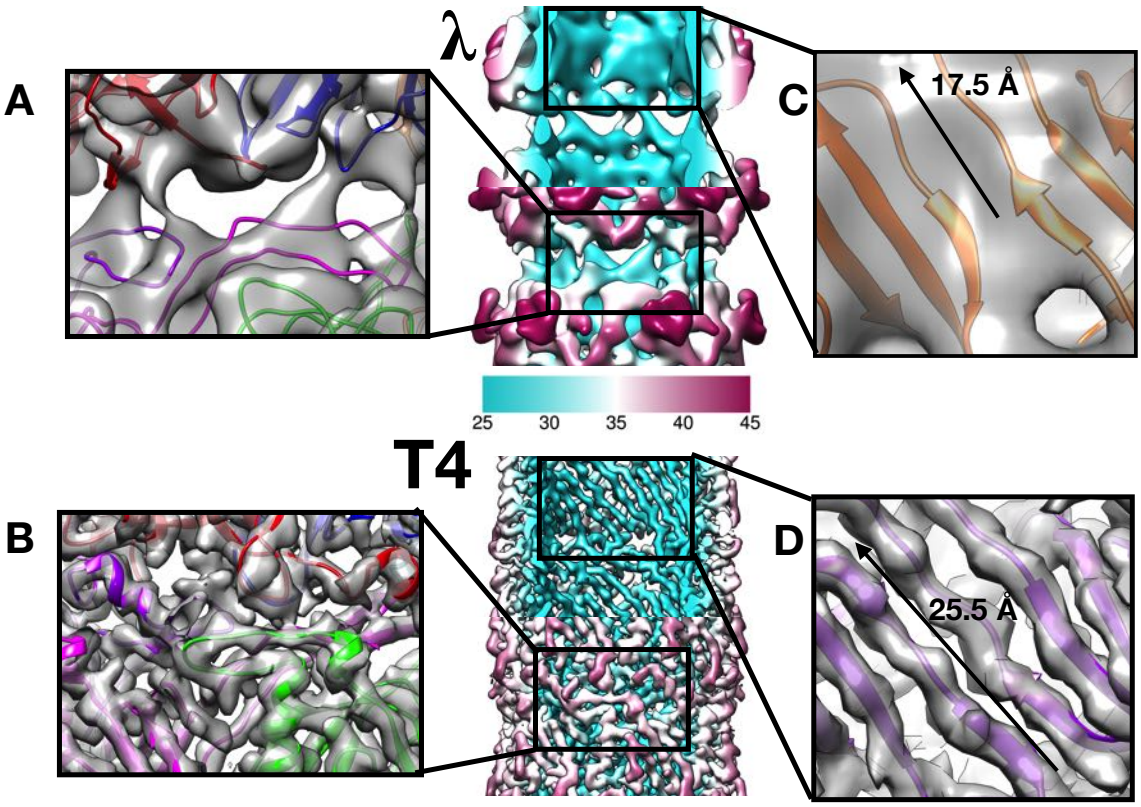

Figure 55. Comparisons of the inner wall and disk connections between λ and T4

A,B) The area between disks is highlighted for λ (A) and T4 (B). C,D) An interior view of the inner wall between two adjoining monomers is highlighted for λ (C) and T4 (D). Scale bar represents cylinder radius in Å.

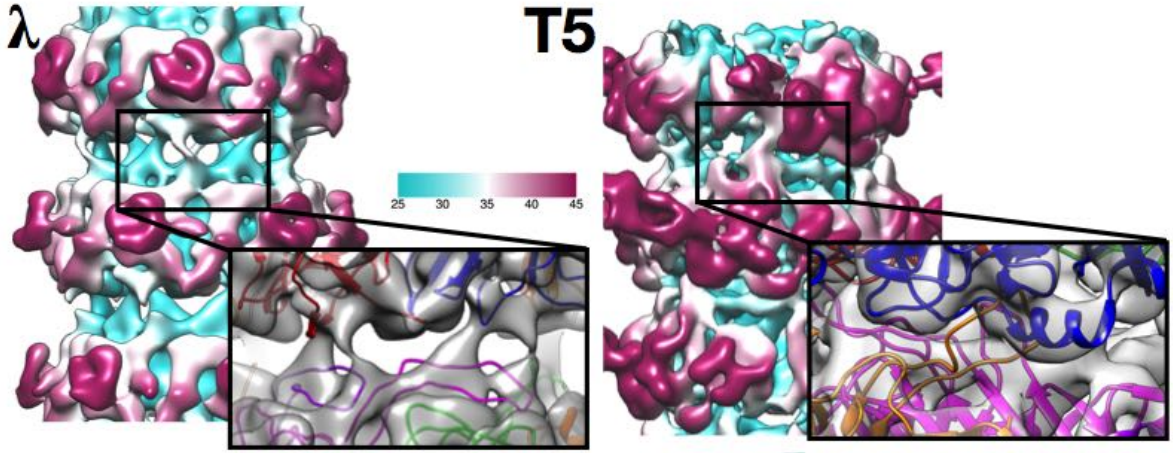

Figure 56. Comparisons of disk overlap between λ and T5

Inserts show the atomic models of λ and T5 in the density maps. Scale bar represents cylinder radius in Å.

λ tail tube polymerization entails considerable structural rearrangement in gpV$_N$ and may be directed by electrostatic interactions

During polymerization, gpV$_N$ undergoes considerable structural rearrangement. This study shows that it is not simply the TE-loop that changes during the polymerization of the tail tube domain but also the termini, the β-sandwich, and the helix. In order for the tmp protein to fit into the tube the termini have to rearrange to the external surface of the tail tube. The TE-loop moves nearly 40 Å away from the β-sandwich (Figure 50) revealing a surface that is negatively charged. It appears that the helix and β sheets of the gpV β-sandwich also reorient relative to each other. This causes the positively charged face opposite the TE-loop to flatten and form into a wedge. Along with the rearrangement of the TE-loop, two oppositely charged surfaces are better exposed to facilitate hexamerization. These rearrangements result in an atomic model that matches that of T4 gp19.

λ tail tube assembly may be directed by electrostatic interactions. Figure 57 provides a schematic for the clustering of charges in individual monomers of gpV and hexameric disks of

gpV. The TE-loop of gpV has a 6 aspartic acid generating a high density of negative charge. T5 does not have this clustering of a single charge in its two enfolding loops (Loop 1: 3 negative and 3 positive charges; Loop 2: 1 positive charge). It is also not seen in the T4 enfolding loop that has 4 negative charges and 3 positive charges. The opposite side from the enfolding loop of the gpV has more positive charges. This suggests that part of assembly requires the N-terminus and TE-loop opening up a negatively charged pocket into which a positively charged surface of an adjoining monomer can fit. The TE-loop then 'reaches over' the positively charged top of the adjoining monomer to make a long continuous surface that the monomers share. Further, the 'bottom' of the disk is positively charged. The negatively charged TE-loop is held in a positively charged pocket generated by the adjoining monomer and the disk above it. The location of the relatively neutral N-terminus remains unknown.

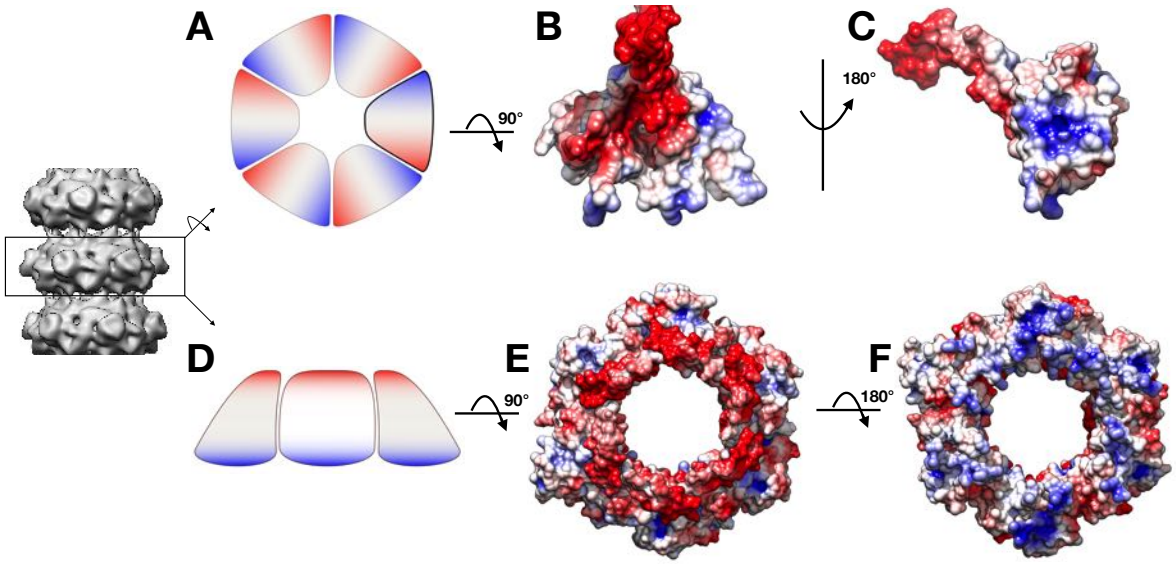

Figure 57. gpV hexamerization and stacking based on electrostatic surface potentials

The color gradient from red to blue represents negative (−5 kT/e) to positive electrostatic potentials (+5 kT/e). Surfaces calculated using APBS from NCBS. A) A cartoon of a top view of an individual hexamer with colors representing charges. B,C) Side views of an individual gpV tube domain monomer. D) A cartoon of a side view of an individual hexamer with colors representing charges. E, F) Top and bottom views of a gpV tube domain hexamer.

The immunoglobin domains may connect end-to-end forming a chain along the exterior surface of the tube.

The Ig-like domain of gpV is the C-terminal 89 residues of the 246 residue length protein that features a canonical Ig-like fold of the I-set class. It is a member of the bacterial Ig-like domain 2 (Big_2) family which is comprised of ~1300 bacterial and 50 phage protein sequences (Fraser *et al.*, 2006; Pell *et al.*, 2010). Several studies have shown that while the tail can still assemble into infectious virion without the Ig-like domain, the efficiency or burst size is reduced (Katsura, 1976b; Pell *et al.*, 2010; Scandella & Arber, 1976). Granted, this depended on the host strain being used. Interestingly, tails without the Ig-like domain were resistant to low-pH dissociation suggesting that the Ig-like domains may introduce some instability into the quaternary structure (Katsura, 1981). Further, to overcome *E. coli pel-* mutants, revertants had mutations in both gpH and gpV (Scandella & Arber, 1976). Both types of mutations appeared to play a role in assembly as well as absorption to the host. It is possible that the Ig-like domain may aid in the efficiency of assembly by stabilizing interactions between disks. The proximal and distal ends of the fold feature negative and positive charges isolated on either end which would promote such associations (Pell et al., 2010). A model of how the Ig-like domains may associate based on electrostatic interactions is shown in Figure 58. The distal and proximal ends of the fold feature highly conserved residues so such a connection may be critical. Notably, defK244 (G222D-Jun Xu unpublished data)(Pell, 2010) which features a mutation resulting in a negatively charged residue in the positively charged proximal patch, has a burst size 50 times less than that of wild-type suggesting that such a change had an effect on folding, assembly, or stability (Katsura, 1976a, 1981). Pseudo-revertants of this mutation, either trimmed a large section of the Ig-like domain off or had a single mutation replacing the proline at 227 with a leucine residue. These also featured burst sizes that were at least

3 times less than that of wild type. In the NMR spectra of gpV_C-G222D and gpV_N, it appears that the mutated Ig-like domain appears to interact with the "top" region of the tube domain as evidenced by chemical shifts in the gpV NMR spectra compared to gpVn alone (Pell, 2010). The region corresponds to the more positively charged area where the negatively charged TE-loop is hypothesized to make contact. Spectra for gpV_C-G222D, gpV_C-G222D/P227L, and gpV_C-P227L suggest that the G222D mutation results in partial unfolding and P227L stabilizes hydrogen bonding between the 7^{th} and 8^{th} β strands thus stabilizing the structure (Pell, 2010). It is interesting to note that the proline at position 227 is supposedly unique to Ig-like domain of gpV (Pell, 2010).

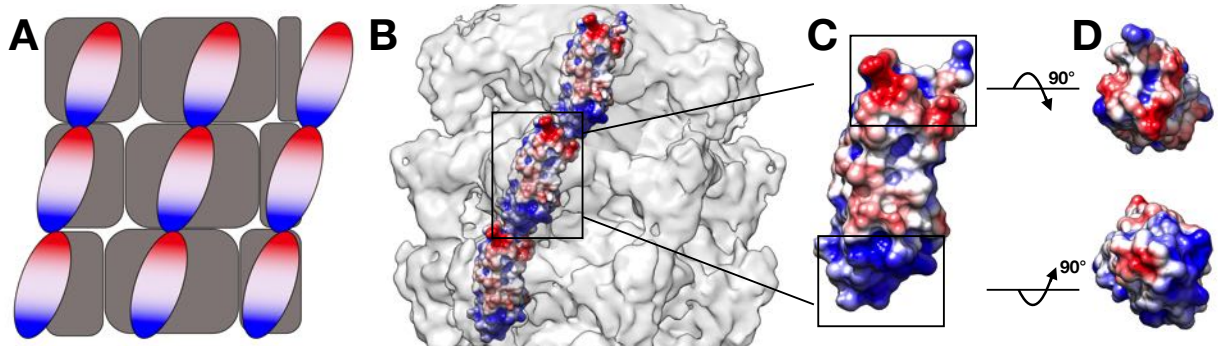

Figure 58. Model of the Ig-like domain orientation and inter-disk connectivity

The color gradient from red to blue represents negative (−5 kT/e) to positive electrostatic potentials (+5 kT/e). Surfaces calculated using APBS from NCBS.A) Cartoon model for how Ig-like domains on stacked disks may connect through charged patches B) Lower threshold map showing Ig-like domain density. Ig-like domains (PDB ID: 2L04) rigid docked into density. C) An individual Ig-like domain surface structure. D) The top and bottom surfaces of the Ig-like domain.

The λ tail tube does not appear change after genome release. It has been proposed for the SPP1 tail that several helices in the SPP1 tail tube protein reorient to trigger DNA release (Plisson *et al.*, 2007). In our structure of λ tails the one helix present does not appear to reorient after ejection. The λ tail structures we presented exhibit minor differences but since they were not all at the same resolution it is hard to determine if these differences are meaningful. The helical

parameters for the three tail samples have similar rise and rotation with differences no greater than 1 Å rise and 1 degree rotation. It has been shown that mutations in gpV that do not affect assembly affected infectivity (Katsura, 1976a; Scandella & Arber, 1976), so gpV may play some active role in DNA injection. Unfortunately, an active role for gpV in DNA injection was not revealed in this study.

The tmp is unstable in the purified tail tube. The tmp protein is known to be present in the tail tube in both *Siphoviridae* (Arnaud *et al.*, 2017) and *Myoviridae* tails (Leiman *et al.*, 2010). In studies in phage λ, anti-gpH serum did not bind to intact phage suggesting that it was mostly buried in the virion (Casjens & Hendrix, 1974). The λ tmp is expected to be in an extended confirmation since it has been reported in contact with the TTC (Roessner & Ihler, 1984) but is cleaved after the addition of gpU to the opposite end of the tail tube (Tsui & Hendrix, 1983). Finally, tail assembly models for phages of *Siphoviridae* and *Myoviridae* involve the extension of the tmp during the polymerization of the tail tube proteins (Abuladze *et al.*, 1994). In this work, a considerable portion of the purified tail tubes appeared hollow in our cryo-EM images. Further, while we did observe density attributable to the tmp in the virion tail tube reconstruction, we did not observe density in the purified tail sample tail tube reconstruction. Together, it suggests that at the very least the tmp is unstable in the purified tail sample tail tube. In U amber mutants, it appears that gpH may 'fall out' of assembled tails and polymerization of gpV will continues polytails (Katsura, 1976b). In our work, mass spectrometry experiments regarding the protein composition of the purified tail samples showed that gpU was present. However, polytails were occasionally observed in the cryo-EM images suggesting some lack or alteration of gpU activity. In the virion tail structure, the tape measure protein density does not appear to contact the tail tube density suggesting that there is limited contact between the two proteins in assembled tails. If there is no

direct contact, the tmp may simply diffuse out of the tail tube over time despite the presence of gpU. Overall this suggests that gpH is not stable in the tail tube.

4.4 Future Directions

What is the structure of the λ tail tube in the absence of the Ig-like domains? An interesting feature of the λ tail tube is the difference in rise compared to structurally similar tubular assemblies in other phages. The λ tail tube features a rise that is nearly 2 Å greater than that of T5 and T4. Also, tail tubes with smaller or absent Ig-like domains were resistant to disassembly in the presence of low pH (Katsura, 1981) suggesting an increased stability. It is also worth noting that compared to T5 and T4 that there is a surprising lack of density between the disks in all three structures for the λ tail tube. If the Ig-like domains do connect end to end, they may play a role in inter-disk connectivity. While they may play this role, the end to end length of the Ig-like domain is slightly greater than the helical rise of the λ tail tube. It's possible that their connectivity introduces a space between the disks resulting in a slightly greater rise than structurally similar systems. It would be interesting to see if the rise of λ tail tube would be more like that of T5 and T4 without the Ig-like domain. Further, if the tail is more stable, a higher resolution map may be obtainable. This would provide further insight into the assembly and structure of the λ tail tube.

The structures of the purified tail tube and the ejected tail tube could not be compared in detail. Unfortunately, the dataset used in this study did not generate a structure which would allow for significant conclusions to be made. In T5, it was assumed that the structure of the purified tail that had been exposed to the host receptor would be similar to that of a full virion that had ejected its genome (Arnaud *et al.*, 2017). It's important to note that the purified tails in this study lost the

tape measure protein without ever being exposed to the host receptor. Since a *Siphoviridae* ejected tail tube has not been studied at high enough resolution, the conclusion that the purified tail tube without the tape measure protein is identical to that of the ejected tail tube is not guaranteed. While a massive reorganization proposed based on work with SPP1 does not appear to be the case, that does not mean that smaller rearrangements don't occur. Obtaining a higher quality dataset of ghost virions would allow for this determination.

5.0 The Phage λ Tail Tip Complex

5.1 Introduction

Phage tail tip complexes (TTC) appear to have great diversity, much like capsids, but the TTC proteins appear to share a common arrangement (Davidson *et al.*, 2012) (Figure 59). The base of the tail tube is capped by a hexameric ring of the distal tip protein (DTP or Dit). Structural similarity between Dit proteins has been found between both Gram-positive and Gram-negative *Siphoviridae*, *Myoviridae*, and bacterial secretion systems (Flayhan *et al.*, 2014; Veesler *et al.*, 2010). Connected to the Dit protein is the basehub protein (BHP) (Davidson *et al.*, 2012) which forms the central structure of the TTC or baseplate. Structural similarity has been found between the BHPs of the T5 (*Siphoviridae*) BHP, pb3, and the T4 (*Myoviridae*) BHP, gp27 and Mu (*Myoviridae*) BHP gp44, (Kondou *et al.*, 2005; Zivanovic *et al.*, 2014). BHPs have 4 distinct domains with domains HDI and HDIII oriented towards the tail tube and domains HDII and HDIV oriented towards the tail tip (Kondou *et al.*, 2005; Sciara *et al.*, 2010; Zivanovic *et al.*, 2014). Attached to the BHP is the receptor binding protein (RBP) that identifies and binds to the host. In phages with a baseplate, there may be several RBPs bound to the Dit and BHP proteins whereas in phages without a baseplate, the RBPs are connected solely to the BHP protein (Mc Grath *et al.*, 2006; Veesler *et al.*, 2010). Structural work on RBPs has focused on the structures of Lactophage RBPs revealing a conservation of domains (shoulder, neck, and head) with differences between phages being concentrated in the neck and head domains (Dowah & Clokie, 2018). Available atomic models are primarily of individual proteins while complete structures of *Siphoviridae* TTCs are less common and of lower resolution.

The λ TTC has been described as 5 disks with a central fiber and is proposed to contain the proteins gpJ, gpK, gpL, and gpI (Figure 60)(Katsura, 1983). The proteins gpM and gpH may also be located at the tail tube-TTC interface. Phage λ also has side tail fibers that are comprised of gp*stf* and require gp*tfa* for proper assembly (Montag & Henning, 1987; Montag *et al.*, 1989). gp*stf* binds OmpC, a host cell receptor, that aids in the efficiency of binding to the host although it is not required for successful infection (Hendrix & Duda, 1992). Phage λ also has side tail fiber genes that would normally produce fibers that are comprised of gp*stf* and gp*tfa*, however, the *stf* gene has a frameshift mutation and only produces a truncated peptide called gpORF 401, which is present in purified phage particles and is thought to be attached to the tail in the region of the TTC. The original isolate of λ, Ur-λ, has the complete genes and has 4 fibers composed of gp*stf* and gp*tfa* (Hendrix & Duda, 1992)(Roger Hendrix and Bob Duda personal communication). Three copies of gpJ form the central tail fiber (Casjens & Hendrix, 1974; Dai, 2009) that is responsible for binding to the *E.coli* protein LamB (Wang *et al.*, 2000). gpL coordinates a $[4Fe\text{-}4S]^{2+}$ cluster which appears to play a role in stabilizing the structure of gpL and potentially the TTC structure (Tam 2013). There may be only one copy of gpI per TTC (Dai, 2009). While the function of gpI is not clear, it was found to be associated with the purified tail and not the virion (Dai, 2009). gpK may be a member of the CHAP (cysteine, histidine-dependent amidohydrolases/peptidases) superfamily (Rigden *et al.*, 2003). gpM is required for assembly and can form a 'pseudo-initiator' with gpL to form polytubes in the absence of gpH (Katsura, 1976b) and there may be 6-8 copies of gpM in the λ virion (Dai, 2009; Katsura & Tsugita, 1977).

This chapter describes efforts to determine the structure and organization of the λ TTC using cryo-EM.

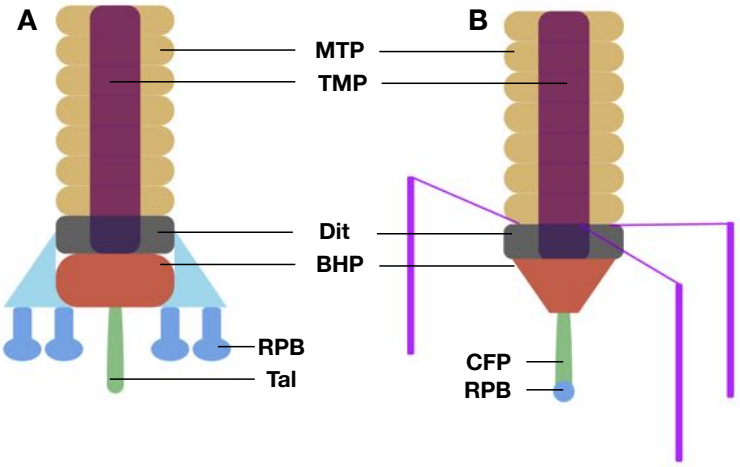

Figure 59. Schematic organization of *Siphoviridae* tail tips

Schematic organization of A) a *Siphoviridae* baseplate structure and B) a *Siphoviridae* tail tip complex are shown. While the arrangement of proteins in tail tips and baseplates may vary, there are several types of proteins that are found in most tails: major tail protein (MTP), tape measure protein (TMP), distal tail protein (DTP or Dit), receptor binding protein (RPB), central fiber protein or tail associated lysin (CFP or Tal). A) A *Siphoviridae* baseplate structure. B) A *Siphoviridae* tail tip complex. This image was adapted from McGrath et al. 2006 and Zivanovic et al. 2014.

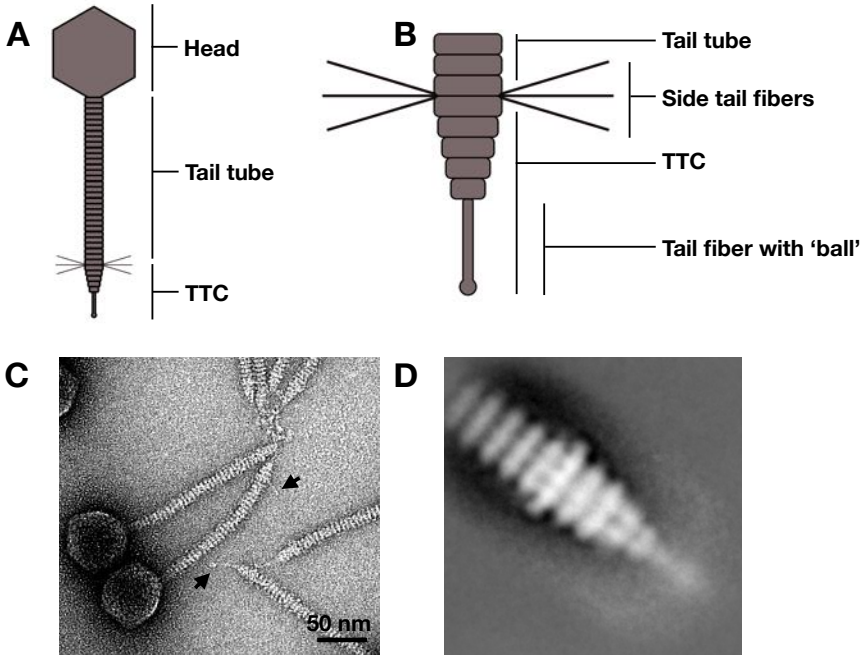

Figure 60. Phage λ tail tip complex

Schematics adapted from Katsura 1983 show: A) the head, tail tube, and TTC are shown in a schematic of the virion, B) a close up of the TTC showing 3 bands of the tail tube, the 5 'bands' that form the TTC, the side tail fibers, and the tail fiber with a 'ball' at the base C) Micrographs showing the TTCs in the virion (Provided by Dr Bob Duda). Arrows indicate the 'ball' shape at the end of the tail fiber as well as a shortened side tail fiber (ORF 401). D) A 2D average of the virion TTC obtained from negative stain images of the virion.

5.2 Results

5.2.1 Imaging of the λ tail tip complex

Phage λ tails were purified as described in Materials and Methods. The widest section of the λ TTC is ~15 nm across, indicating that very thin ice would be optimal for obtaining good contrast of the TTC structure. We tried several approaches to achieve good contrast (Figure 61). The TTC had an affinity for the carbon grid, resulting in poor distribution into the holes of the cryo-grids. To counteract this tendency, a Quantifoil grid with a 2nm thin layer of continuous carbon was used to encourage better distribution of particles. In addition, the carbon layer aids in estimating defocus by providing strong Thon rings in the power spectra. However, the carbon layer also reduces the contrast of the TTCs, affecting orientation determination and the signal-to-noise ratio in reconstructions., Cryo-negative staining was also used to increase contrast (see Chapter 2 for method). A cryo-negative stain dataset obtained using a Quantifoil grid showed better contrast than the previous datasets, but the number of stable TTCs was low. A second dataset using C-flat grids that were reported to better for thinner particles and ice (Cho *et al.*, 2013) resulted in sufficient contrast as well as a useful distribution of particles.

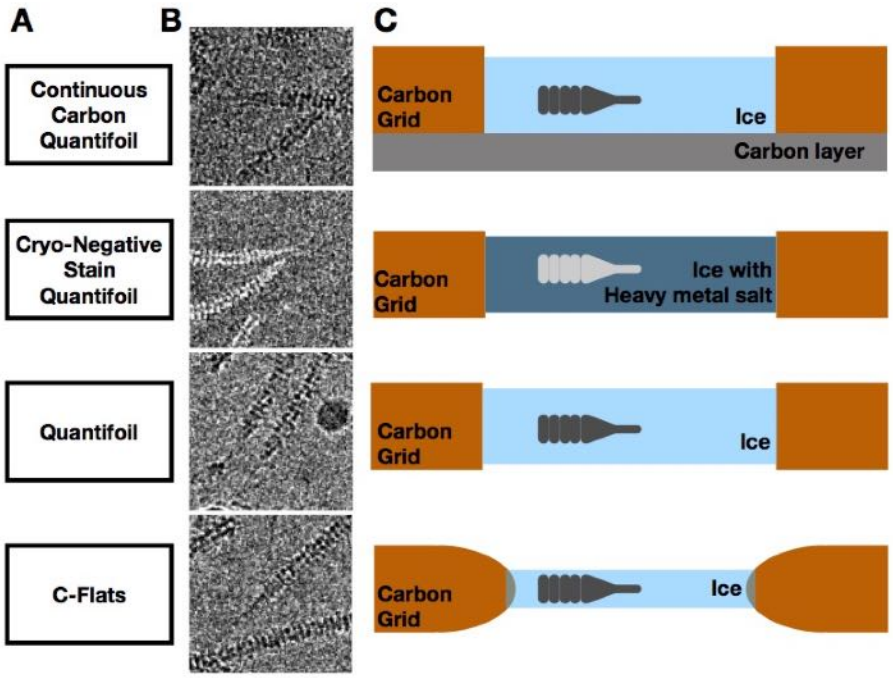

Figure 61. Imaging of the λ TTC

A) The conditions of the imaging datasets. B) Micrographs representing the most ideal TTC found. C) Schematics of the imaging conditions.

5.2.2 The purified tail TTC shows a high degree of variability and 'dots' in the virion TTC

Negative-stain EM images of the λ virion TTC has previously revealed a conical tip comprised of ~5 disks with a central fiber (Katsura, 1983). Our EM imaging of the purified tail TTC revealed a variety of structures (Figure 62), including many that had the expected conical tail tip. Abnormal TTCs appeared to lack half of the conical shape, much like a fountain pen viewed from the side, or appeared shorter and blunt, or appeared to have lost the conical end altogether and instead had two strands extended from the tail tube into a 'crab claw'. The cryo-negative stain EM dataset also revealed particles deviating from the conical tip shape (Figure 62), however, ~30% of TTC particles had the expected conical shape. In some cases, the stain appeared to form a

continuous channel connecting the tail tube to the external environment (discussed in Chapter 4). In all datasets, only a few thousand tail tip particles of good quality were obtained.

A dataset of virion tips was also obtained to both study the tail tube and TTC. The virion TTC also had a high affinity for the carbon grid with few entering the holes. This was remedied by using a C-flat grid and double-blotting prior to plunge freezing. Images of the virion TTC revealed zero to four electron dense 'dots' around the widest point of the tail tip (Figure 63). These dots could be attributed to either the shortened tail fibers (Chapter 6) oriented perpendicular to the camera resulting in the projection of the 'stubs' of side tail fibers composed of gpORF 401 as 'dots', or to the strongly scattering Fe-S clusters in gpL (Tam *et al.*, 2013). The intensity of the 'dots' is variable, possibly due to different defocus settings used during data collection. If the 'dots' are actually the gpL Fe-S clusters than this is a novel observation. The 'dots' were not observed in the purified purified tail samples which do not have the *stf* and *tfa* genes as they are not present on the plasmid used to express tails. The plasmid used for making purified tails does not carry the aberrant side tail fiber gene that encode the gpORF 401 thought to make the tail fiber stubs, which correlates with its lack of 'dots'. This strongly suggests the 'dots' are more likely to be due to the orientation of the shortened side tail fibers.

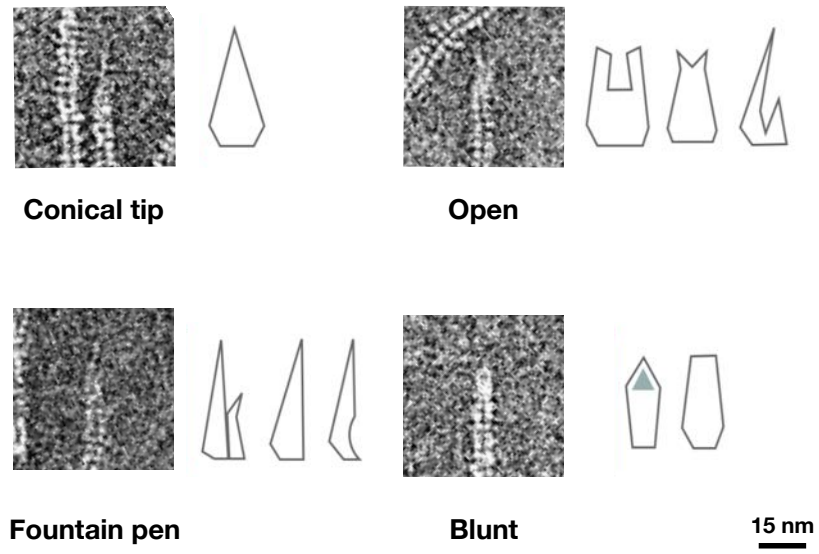

Figure 62. Variations in the purified tail TTC morphology observed in the cryo-negative stain experiment

Ideal particles have the conical tip similar to that detailed in Katsura 1983. Other shapes are named based on how they appear such as 'fountain pen', 'blunt', and 'crab claw'. Schematics next to the images show the variation in each of the alternate types of TTC shapes found.

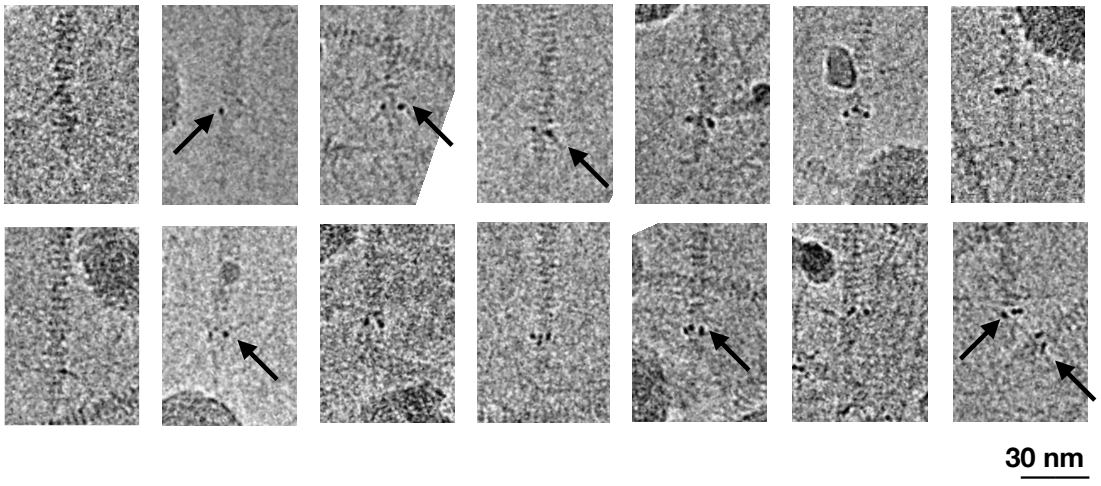

Figure 63. 'Dots' observed in the virion TTC by cryo-EM

This gallery of virion TTC show the variation in the number of 'dots' observed both in the number and intensity observed.

5.2.3 2D class averaging suggests 3-fold symmetry in the λ TTC

TTCs that appeared to match the canonical conical λ TTC were selected for reconstruction, but the variable TTC structures were not included. 2D class averages performed using EMAN2 of both the purified tail and virion TTCs were calculated to aid in generation of a starting model for 3D structure analysis as well as to determine if there were distinct classes of tail tips that were not discernible by eye. 2D averages of the purified tail TTC show density protruding from one side of the tip at the narrow end of the TTC (Figure 64). In some 2D averages, weaker density is present on both sides of the TTC with stronger density present in the center. This additional density appears to protrude from the 'band' that connects to the tail fiber and suggest a 3-fold symmetric structure.

2D class averaging of the virion TTCs tended to be dominated by the 'dots', so a filter was applied to remove their influence on particle orientation. The resulting 2D averages showed that the 'dots' are located at the tail tube-TTC interface (Figure 65). Class averages of the virion tail tip also appear to have density protruding from the side of the TTC at the narrow end of the tip (Figure 65). Similar to the protruding density in the purified tail samples, the extra density is protruding from the 'band' that connects to the tail fiber. Comparing the 2D averages of the purified tail (Figure 6) to the 2D averages of the virion show the attachment point of the side tail fiber stubs (the presence of which is confirmed in Chapter 6). In the purified tail, the base of the conical tip appears to be narrower than the tail tube but in the virion is wider than the tail tube. Since the *stf* and *tfa* genes are not present on the plasmid used to make purified tails, we conclude that this extra width corresponds to where gpORF 401, the short tail fiber stubs, connect to the TTC. Also, there appears to be some density extending from the TTC which may be side tail fibers.

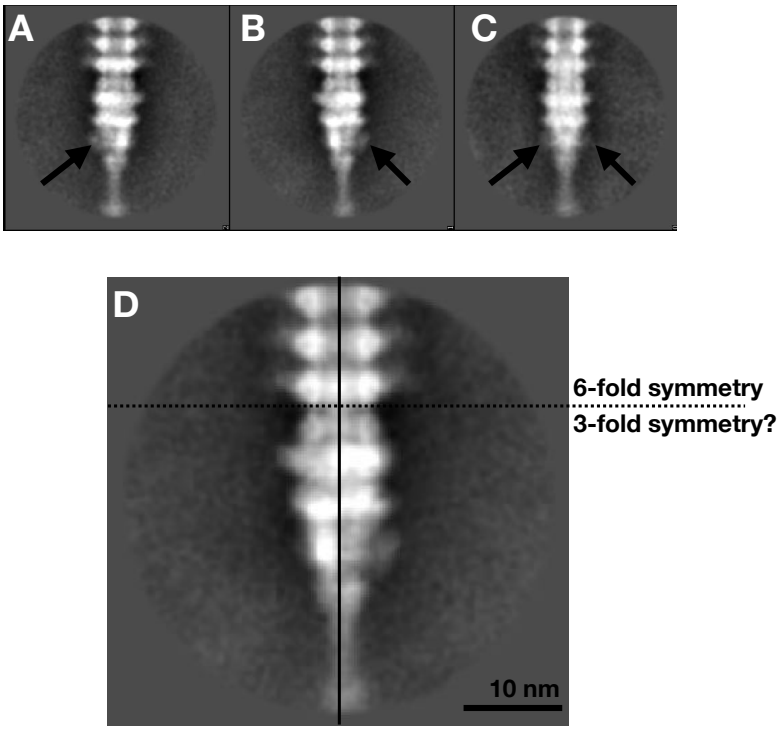

Figure 64. 2D class averages of the purified tail TTC obtained from cryo-EM experiments

A-C) 2D class averages are shown with arrows pointing to density protruding from the narrow end of the TTC. B) A schematic showing the possible symmetries highlighted on the 2D average.

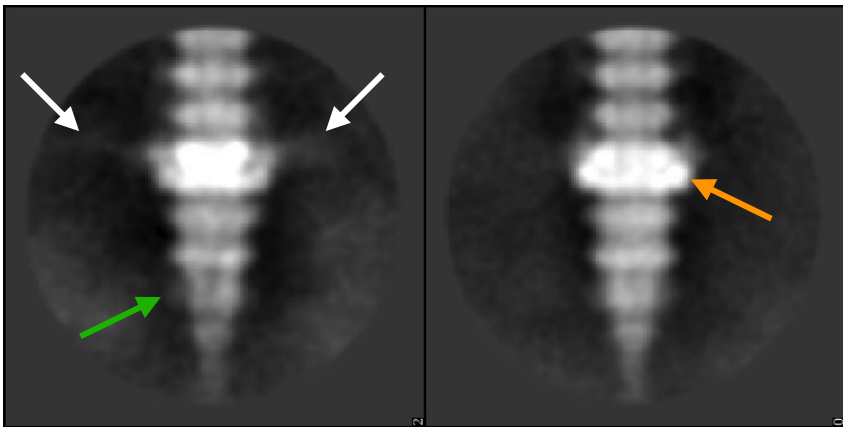

Figure 65. 2D class averages of the virion TTC

Class averages reveal strong density that may be the 'dots' observed in EM images (orange arrow). Density protruding from the side of the TTC that was observed in the plasmid tail TTC is also present in the virion TTC 2D averages (green arrow). Density is protruding from the top of the TTC (white arrows).

5.2.4 The structure of the λ TTC

We have generated a ~8 Å map of the λ plasmid tail TTC. This is one of the first higher resolution structures of a Gram-negative feeding bacteriophage *Siphoviridae* TTC reported.

5.2.4.1 A subnanometer EM density map of the λ TTC

3-fold symmetry is found in the baseplates of both *Myoviridae* (Taylor 2016) and *Siphoviridae* (Sciara *et al.*, 2010). Applying 3-fold symmetry to reconstructions of the λ TTC resulted in convergence in the tail tube sections of the structure as well as a sizeable portion of the TTC. The map appears to have a resolution of ~8 to 9 Å. The three 6-fold symmetric tail tube disks have recognizable features such as the O-loop of the tail tube domain and the Ig-like domains (discussed below). The top band of density of the TTC is narrower-than both the tail tube above and the TTC below, as seen in the EM images and the 2D averages. The outer surface of the TTC features several 'bumps'. At the widest section there are two rows with 6 distinct bumps that are not equal in size or distributed at equal angles like the features of the tail tube disks. As the TTC narrows towards the central tail fiber, there are 3 very distinct bumps separated by relatively featureless density. Rotating the TTC density shows that the additional density protruding from the side of the TTC observed and described in the 2D averages corresponds to the 3 distinct bumps (Figure 66). The central tail fiber has a similar shape to the fibers observed in EM images but it is poorly defined demonstrating its relative flexibility. Overall, the TTC density matches the 2D averages suggesting that applying 3-fold symmetry is appropriate.

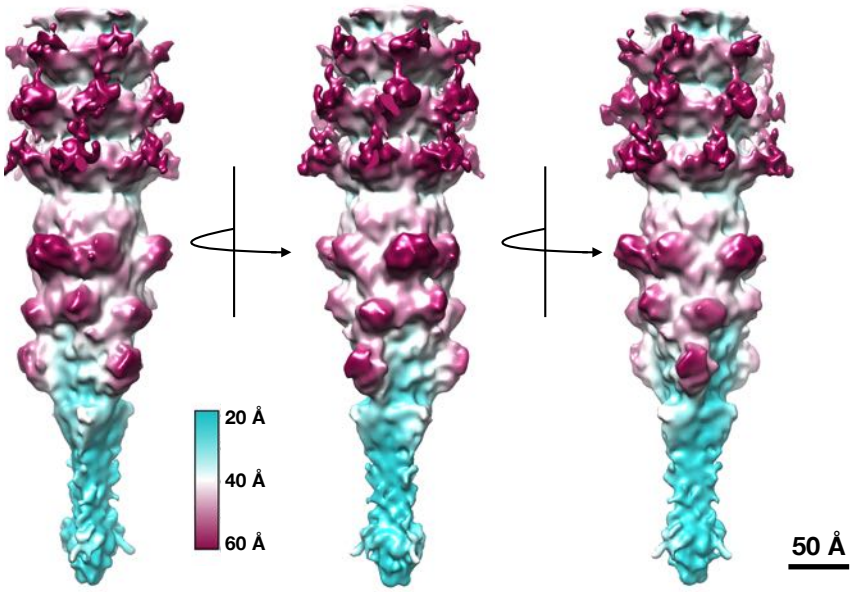

Figure 66. The plasmid tail TTC seen from different angles

5.2.4.2 The virion has a pseudo-baseplate structure not present in the plasmid tail

As stated earlier, the genes *stf* and *tfa* were not present in the plasmid used to produce tails. The truncated form of protein gp*stf*, gpORF 401, is present in the virion TTC (Chapter 6). Comparing the density maps for the plasmid tail and the virion show where these proteins may be located and provide some initial insight into how they are connected to the TTC. While the resolution of the virion TTC is low, there is extra density at the tail tube-TTC interface (Figure 67, arrows). This extra density is also observable in the 2D averages of the virion TTC when compared to the 2D averages of the plasmid tail TTCs (Figure 64,65). Overall this suggests that the λ side tail fibers connect at the tail tube-TTC interface forming a pseudo-baseplate like structure.

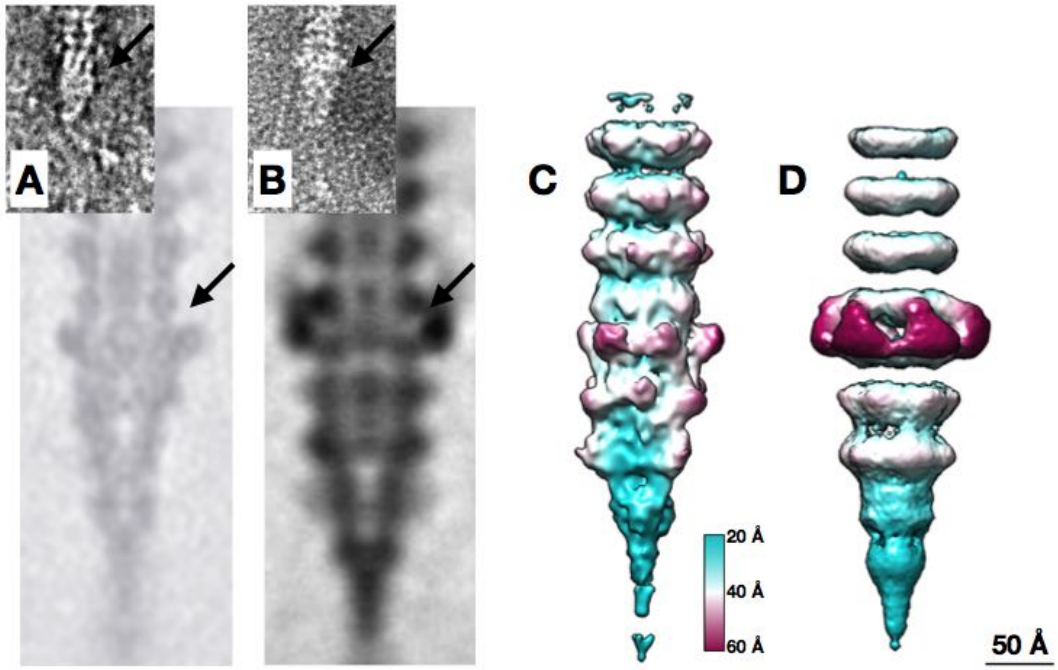

Figure 67. The λ TTC may have a pseudo-baseplate

A) An EM image of the plasmid tail TTC is shown with a cross section of the plasmid tail TTC reconstruction. B) An EM image of the virion TTC is shown with a cross section of the virion TTC reconstruction. C,D) The reconstruction of the plasmid tail TTC is shown next to the reconstruction of the virion TTC. The density maps are colored based on distance.

5.2.4.3 The orientation of the tail tube protein is discernable

Several disks of the tail tube were included in particle images, because we expected any meaningful reconstruction would successfully include recognizable tail tube features in the density. The density associated with the tail tube proteins are resolved enough to decipher the orientation of the tail tube protein relative to the TTC at higher threshold values (Figure 68). The distinctive Ig-like domain densities are visible and show the same Ig-like domain connectivity observed in the helical reconstructions of the tail tube (Chapter 4). At lower thresholds, features of the tail tube domain are discernible. In particular, the O-loop is identifiable in the density. The density suggests that the TE-loop is oriented towards the TTC. In the baseplate structures for the

T4 (*Myoviridae*) and the p2 (*Siphoviridae*), the TE-loop of the tail tube domain is also oriented towards the structure that makes contact with the host.

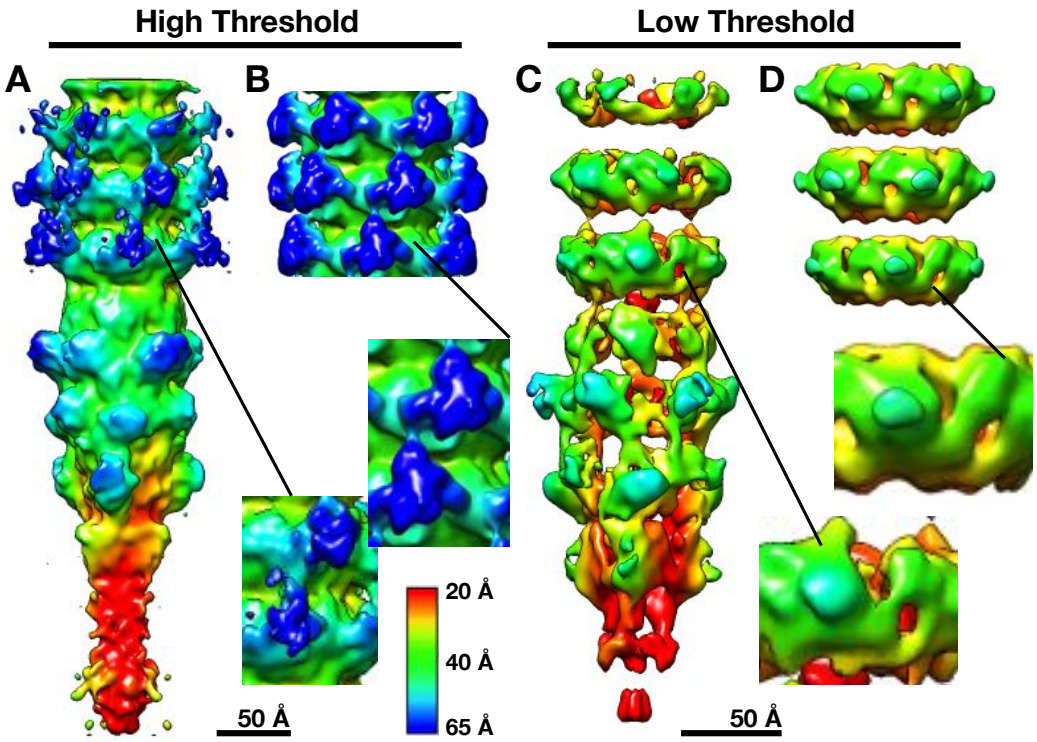

Figure 68. The orientation of the major tail tube protein, gpV, is discernable in the plasmid tail TTC

A) The plasmid tail TTC reconstruction is shown at a higher density threshold so that the Ig-like domain densities are visible. B) The plasmid tail tube reconstruction is shown at a higher density threshold so that the Ig-like domain densities are visible. C) Inserts show the Ig-like domain densities. D) The plasmid tail TTC reconstruction is shown at a lower density threshold so that the tail tube density is visible. E) The plasmid tail tube reconstruction is shown so that the tail tube density is visible. F) Inserts show the tail tube density in greater detail.

5.2.4.4 Cross sections of the TTC density map reveal a 'plug'

Cross sections of the purified tail tube did not show internal density suggesting that the tape measure protein was not present (Chapter 4). Cross sections of the plasmid tail TTC reconstruction also show a hollow tail tube but reveal a 'plug' density that fills the majority of the TTC density (Figure 69). In the plasmid tail TTC structure, the density associated with the 'plug' extends to the first hexamer of the tail tube. While the plug appears to make no contact with the

tail tube density it appears to make limited contact with the TTC. The plug density does not fill the entirety of the TTC density as there is an additional hollow section located at the very end of the tip. In the lower resolution structure of the virion TTC, the density of the plug and the density in the tail tube is continuous. The identity of the protein or proteins that comprise this density is unknown.

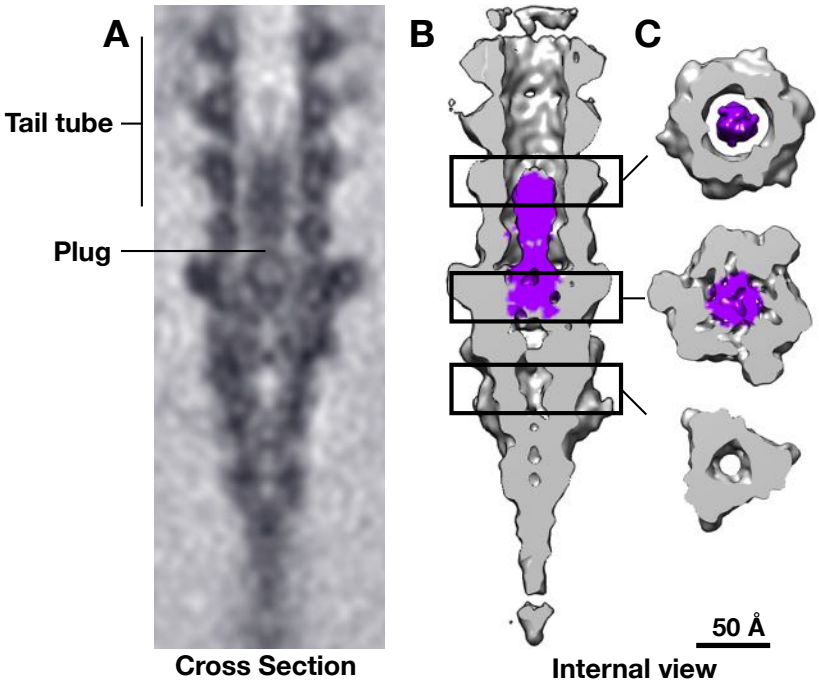

Figure 69. Cross sections of the TTC density map reveal a 'plug'

A) A cross section of the plasmid tail TTC reconstruction is shown with the tail tube and plug highlighted. B) An internal view of the plasmid tail TTC reconstruction is shown with the plug highlighted in purple. C) Top view of the plasmid tail TTC reconstruction are shown with the plug highlighted in purple.

5.2.4.5 The λ TTC may share structural similarities with other phages

The λ TTC appears to *have* 3-fold symmetr*y* like the inner baseplate hubs observed in *Myoviridae*. When comparing the plasmid tail TTC to the inner baseplate structure of T4 (*Myoviridae*), a structurally similar conical shape is visible (Figure 70). It is possible to fit the atomic model for the T4 inner baseplate into the plasmid tail TTC density since they also share

similar dimensions. The T4 baseplate central spike protein, gp5, has a lysozyme domain that protrudes from the central spike. The lysozyme domain of T4 fits neatly into the 3 distinct bumps that protrude from the narrow section of the TTC. Further, the location and identity of the T4 proteins after fitting them into the λ TTC density may provide some insight into the location of the λ proteins (Figure 71).

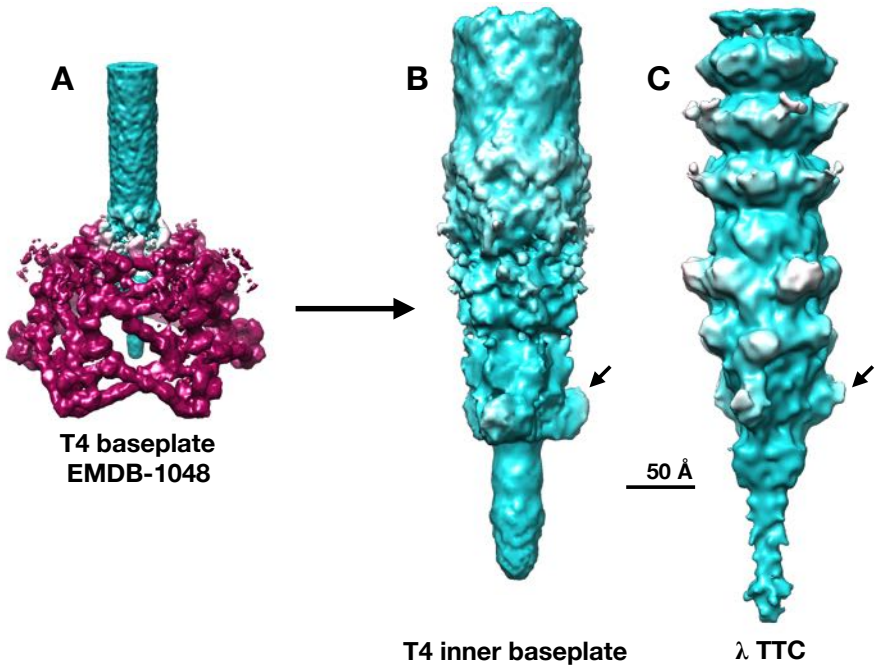

Figure 70. The inner baseplate and central spike protein complex of T4 is similar to the λ TTC density map

A) The density map of the T4 baseplate and tail tube is shown. B) The tail tube, inner baseplate, and central spike density of T4 is shown without the baseplate visible. C) The plasmid tail λ TTC is shown with arrows highlighting specific features that are similar between T4 and λ.

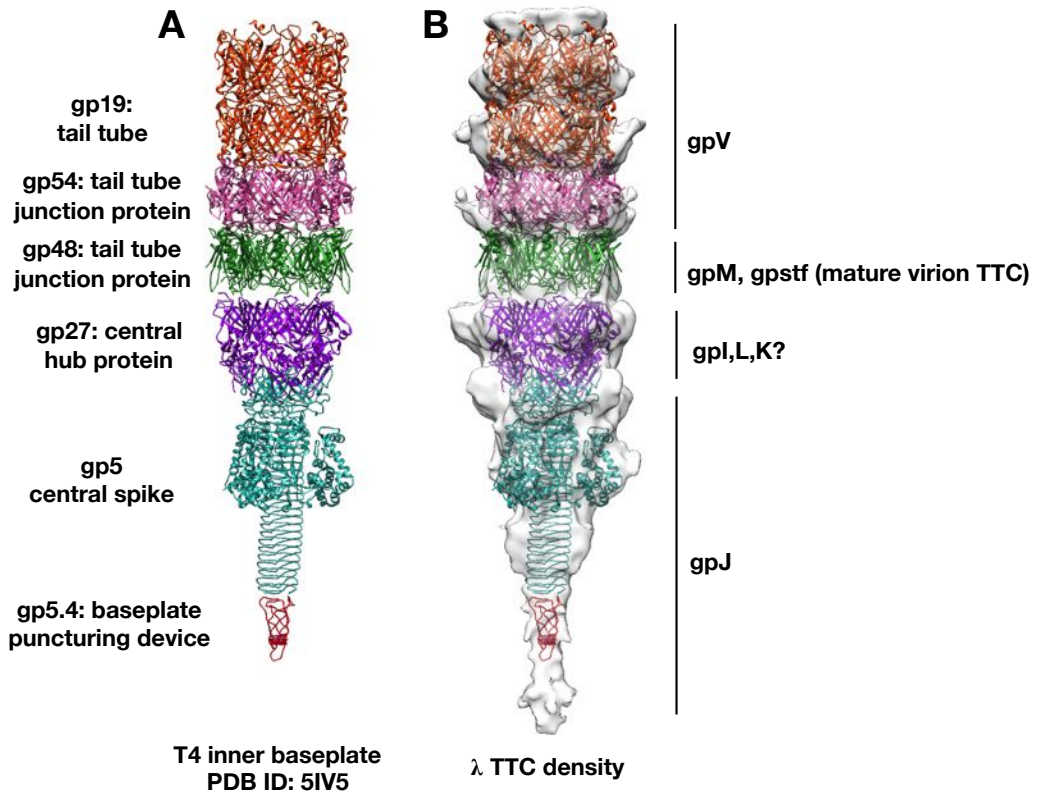

Figure 71. Fitting the atomic model for the inner baseplate and central spike protein complex into the λ TTC

A) The atomic model for the tail tube, inner baseplate and central spike complex of T4 is shown with the identity of the individual proteins identified. B) The atomic model shown in A is fitted into the λ TTC density. The potential locations of λ TTC proteins are identified.

5.3 Discussion

This work provides evidence that the λ TTC features 3-fold symmetry, while the tail tube has 6-fold symmetry. Phage tails often feature 6-fold or pseudo 6-fold symmetry. Additionally, 6-fold symmetry has been reported in both *Myoviridae* and *Siphoviridae* baseplates. 6-fold symmetry has been well described in the baseplate structure of the *Myoviridae* phage T4 (Kostyuchenko *et al.*, 2003; Taylor *et al.*, 2016). *Siphoviridae* baseplate structures have also been reported to have 6-fold symmetry, including those from phages 1358, phi11, p2, TP901-1, and Tuc2009 (Bebeacua *et al.*, 2010; Koç *et al.*, 2016; Legrand *et al.*, 2016; Sciara *et al.*, 2010; Spinelli *et al.*, 2014). A

transition between the 6-fold symmetry of the tail tube and 3-fold symmetry in part of the baseplate or TTC has also been described. The central spike complex that comprises the inner baseplate hub of *Myoviridae* T4 features 3-fold symmetry (Kanamaru *et al.*, 2002; Taylor *et al.*, 2016). In the *Siphoviridae* phage p2 baseplate, the BHP protein, ORF16, forms a trimeric dome that closes the tail tube off from the external environment (Sciara *et al.*, 2010). In phages with smaller tail tip structures such as SPP1, elements of 3-fold and 6-fold symmetry have been reported (Goulet *et al.*, 2011; Veesler *et al.*, 2010). Structural similarities in the BHPs of both *Myoviridae* and *Siphoviridae* suggest that the HD IV and HD I domains allow for the 6 fold symmetry to 3 fold symmetry transition between the tail tube and any tail fiber or central spike complex (Zivanovic *et al.*, 2014). Strong evidence suggests that 3 copies of gpJ are present in the λ TTC (Casjens & Hendrix, 1974; Dai, 2009). Taking this and previously reported symmetry in other *Siphoviridae* tail tip structures into consideration, 3-fold symmetry was used to generate a TTC density map. Identifiable features in the tail tube section of the structure and similarities with the 2D averages strongly suggest that this structure is relevant. Finally, this means that the λ TTC has a similar arrangement of proteins as previously reported baseplates and TTCs of both *Myoviridae* and *Siphoviridae* phages.

This work allows the formation of a model regarding the locations of the tail tip proteins (Figure 72). Fitting the T4 inner baseplate and central spike complex atomic model into the TTC density suggests general regions for the locations of proteins. The difference in diameter at the tail tube-TTC interface between the purified tail and virion TTC 2D class averages and density maps illustrates where gp*stf* likely binds to the TTC. The plasmid used to express tails did not include gp*stf* or gp*tfa* whereas the virion did include these proteins (Chapter 6) so the change in diameter between the two reconstructions most likely corresponds to the presence of the truncated side tail

fibers. Further, we now know that it is the N-terminal gpORF 401 of gp*stf* that is bound to the TTC (Haggård-Ljungquist *et al.*, 1992)(This work-chapter 6). gpM does share sequence similarity with the Siphoviridae phage T5 Dit protein, pb9, which connects to the T5 L-shaped fibers (Flayhan *et al.*, 2014; Zivanovic *et al.*, 2014). The structurally similar Dit proteins pb9 and *Siphoviridae* phage SPP1's gp19.1 are both hexamers justifying the application of 3-fold symmetry to the TTC reconstruction. However, 4 side tails fibers as opposed to 6 have been observed on the Ur-λ TTC (Hendrix & Duda, 1992)(Dr Robert Duda and Dr Alexis Huet personal communication). How gpM connects to the side tail fibers is currently not clear. Efforts to apply 2-fold and no symmetry to the reconstruction of the λ plasmid tail TTC did not help resolve this question as the resolution was too low and the side tail fibers are not present. Additionally, the virion TTC reconstruction is also at too low of a resolution to provide insight. BHPs have 4 distinct domains with domains HDI and HDIII oriented towards the tail tube and domains HDII and HDIV oriented towards the tail tip (Kondou *et al.*, 2005; Sciara *et al.*, 2010; Zivanovic *et al.*, 2014). gpL may share structural similarity with the HDI domain (Zivanovic *et al.*, 2014) of the *Siphoviridae* T5 TTC BHP, pb3 which would suggest that gpL is oriented towards the tail tube. gpJ may potentially comprise the majority of the conical tip and the tail fiber (Dai, 2009; Hohn & Katsura, 1977; Kellenberger & Edgar, 1971). A region of gpJ may share structural similarity with the HDIII/IV domains of the T5 TTC BHP pb3 (Zivanovic *et al.*, 2014). The HDIII domain is oriented towards the tail tube and the HDIV is oriented towards the tail tip and tail fiber (Kondou *et al.*, 2005; Sciara *et al.*, 2010; Zivanovic *et al.*, 2014). Thus, gpL and gpI may be located at the top of the TTC connected to gpM and gpJ may comprise the rest of the TTC and central tail fiber.

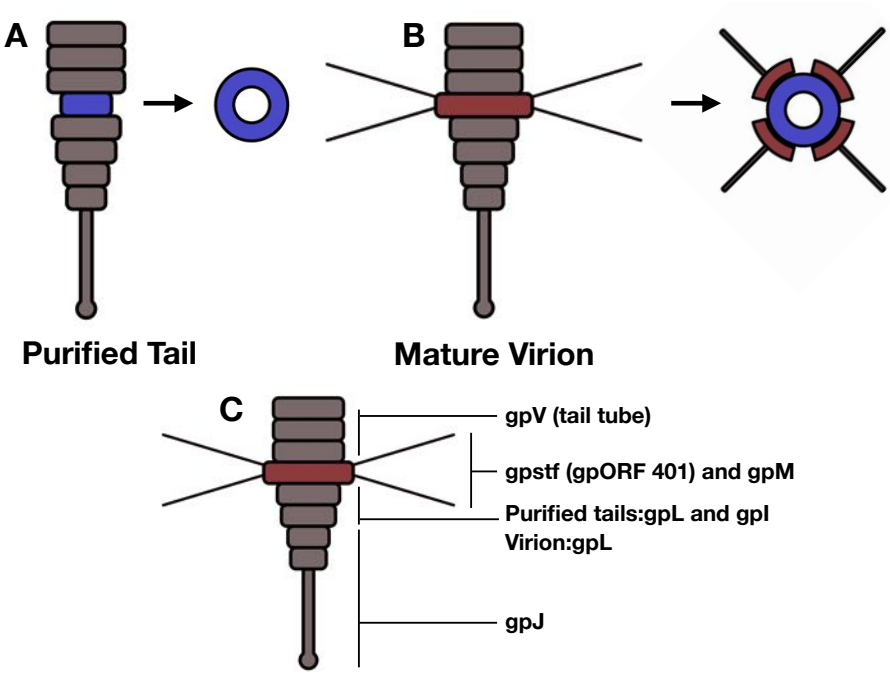

Figure 72. Schematics of the purified tail and virion TTCs

A) A schematic of the purified tail TTC with the model for the tail tube-TTC interface highlighted. B) A schematic of the virion TTC with the model for the tail tube-TTC interface highlighted. C) A model that shows the potential locations of proteins in the purified tail and virion TTCs as well as the tail tube.

This work also suggests that the λ TTC may be very flexible, as indicated most clearly in the cryo-negative stain dataset. The images suggested that either protein being lost from the tail tip or the tail tip was opening. gpJ has been shown to be lost due to osmotic shock (Konopa & Taylor, 1979) and the cryo-negative stain grid preparation may have been more destabilizing than the other approaches. In addition, both the purified tail sample and virion purifications featured steps that utilized high salt that could also have destabilized the tail tips, reducing the count of intact particles for structural characterization. Further, as demonstrated in Chapter 4, the tmp protein gpH is unstable with the appearance of hollow tail tubes and a lack of density within the purifid tail tube reconstruction. The tmp protein is believed to comprise a major portion in other *Siphoviridae* tail tips (Bebeacua *et al.*, 2010; Mc Grath *et al.*, 2006) and a 3 kDa portion of gpH* appears to be partially or completely exposed to the environment when gpJ is lost to protease

degradation (Roessner & Ihler, 1984). If the tmp comprises a portion of the λ TTC, the instability of the tmp may make structurally characterizing the TTC problematic, especially as I did not differentiate between hollow and full tail tubes when selecting TTC particles for analysis. It is important to note that the protruding density at the narrow end of the TTC tip was similar for both the virion and the purified tail, which did not have hollow tubes, suggesting that the mix of hollow and full tail tubes for the virion did not affect TTC structural analysis.

A major hurdle in this study was obtaining sufficient cryo-EM images of high-quality particles. I was able to collect only a few thousand images of TTC particles that were of good quality. Nonetheless, the ~10 Å TTC model achieved for the purified tail sample demonstrates the quality of images collected on the Polara and Falcon 2 camera and offers good hope for improving the structure further with a reasonable increase in the dataset size.

5.4 Future Directions

There are currently no structures of Gram-negative *Siphoviridae* TTCs better than the subnanometer model developed here. Further improvements in the structure would improve understanding of evolutionary relationships among phages as well as of phage-host interactions. Several strategies could be explored to extend resolution towards atomic resolution, including increasing the number of TTC particle images collected as noted above, and to follow the random conical tilt (RCT) method for generating a starting model (Sorzano *et al.*, 2015). The low particle counts resulted partly from the occupancy of ice by the long tail tube, and truncating these (Katsura & Hendrix, 1984) would increase the effective concentration of TTCs in the images. At the extreme, assembling and purifying a complex comprised solely of the TTC proteins is possible

(Dai, 2009), and would maximize the density of TTC complexes in the cryo-EM images, but may not yield authentic TTC structures. Of course, simply collecting more images will also expand the dataset. The RCT method offers greater orientation sampling of particles such as the TTC that may be constrained by the thin ice to be always oriented perpendicular to the electron beam.

6.0 Phage λ Tail Assembly

6.1 Introduction

Phage λ tail assembly (Figure 73) requires the proteins encoded by phage genes Z, U, V, G, T, H, M, L, K, I, and J and can be divided into three general stages: 1) the formation of the initiator 25S complex; 2) the polymerization the tail tube onto the initiator; and 3) termination of polymerization and head attachment. The formation of the 25S initiator complex formation requires the joining of gpM with the 15s tail-tip complex and the tape measure protein coated with chaperones gpG and gpGT. The major tail tube protein, gpV, polymerizes on the initiator while displacing the tmp chaperone proteins. Polymerization pauses once the length of the tail tube reaches the length of the tmp. The tail terminator protein, gpU, terminates gpV polymerization. Concurrently, the tmp protein, gpH, is cleaved, resulting in gpH* (Murialdo & Siminovitch, 1972; Tsui & Hendrix, 1983) and the addition of gpZ 'activates' the tail for head attachment which occurs last.

This chapter focuses on the work to characterize the initiator complex and the fate of tail proteins during tail assembly. The first section of this chapter describes our work to characterize the initiator. The second section compares the protein components present in the initiator, to those present in purified tails and virion tails, to confirm whether the minor tail proteins form a part of the structure or are only required during assembly.

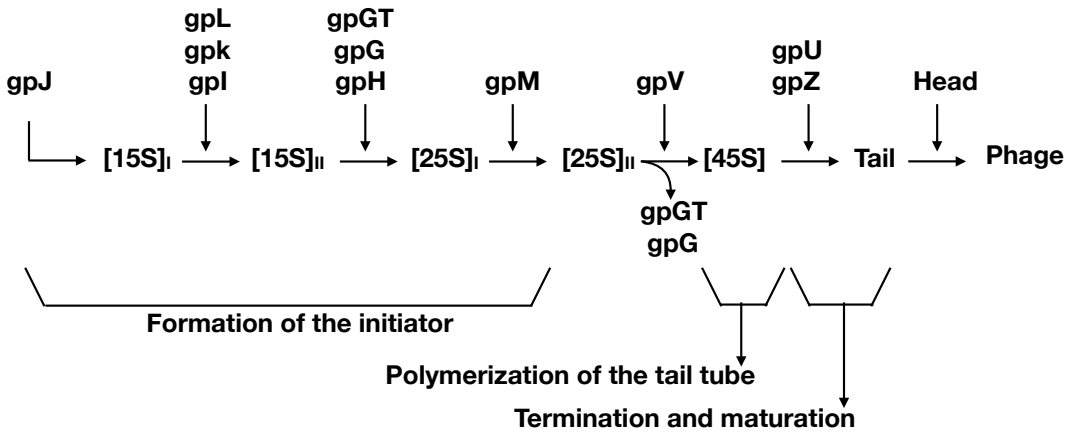

Figure 73. The λ tail assembly pathway

The TTC protein gpJ forms the $15S_I$ complex. gpJ, gpK, and gpI join with the $15S_I$ complex to form the $15S_{II}$ complex. The tmp protein and its chaperones, gpG and gpGT, join with the $15S_{II}$ complex to form the $25S_I$ complex. The initiator is formed with gpM is added to the $25S_I$ complex to form the $25S_{II}$ complex. The major tail tube protein polymerizes onto the initiator while displacing gpG and gpGT. Once the tail tube reaches the full length of the tmp, the tail terminator protein, gpU, stops gpV polymerization. The tail is 'activated' by the addition of gpZ. The tail and head are joined together by the tail attached to the portal containing vertex on the capsid to form a virion. This image was adapted and updated from Katsura and Kuhl 1975.

6.2 Results

6.2.1 Identification and characterization of a tail initiator-like complex

The $15S_{II}$ and $25S_I$ complexes shown in Figure 73 were identified in experiments using in vitro complementation and sedimentation analysis of crude extracts of mutant-infected cells (Kuhl & Katsura, 1975). However, a $25S_{II}$ fraction that contained mostly host proteins (from an amU-phage-infected cell lysate) is the only sample in which a structure that resembles the TTC has ever been observed using electron microscopy.

We set out to purify the initiator by expressing all the proteins expected to form the $25S_{II}$ complex with a N-terminal polyhistidine tagged gpG. Since it was hypothesized that gpG is abundant in the initiator complex (Xu, 2001), a his-tagged gpG was included to provide a handle

for purification by nickel-affinity chromatography. A cell lysate containing the initiator proteins was applied to a Ni column, washed, and eluted with imidazol. The proteins gpGT, gpH, gpJ, and others eluted from the column with gpG (Figure 74A). EM images taken of the eluted peak fraction showed long, thick, continuous wavy structures along with many single U-shaped segments of varying thickness (Figure 74B).

While the imidazol-eluted fractions contained the initiator complex proteins, other additional proteins were present. Size-exclusion chromatography using Sephadex-400 was used to further purify this fraction (Figure 75A) and the bulk of the initiator complex proteins were found in the void volume. While SDS gel analysis showed that more individual bands were present than expected, only tail proteins were identified when the same samples were analyzed by mass-spectrometry. Thus, many of the 'extra bands' appear to be degradation products of the expected proteins and not host contaminants. The EM images taken of the void volume-fractions showed again distinct U-shaped particles (Figure 75B). The single U-shaped particles are generally about ~10 nm thick at the base of the 'U' and have a length of ~35 nm from tip to tip of the 'U' (Figure 76). A majority of the particles were 'U' shaped, but an apparent flexibility in the particle resulted in a scrunched up 'U' shape in some particles. End to end, the whole structure was ~45 nm. The two ends of the U structure are different with one end narrower than the other (Figure 76). In many particles a ball-like shape appears to be attached to the narrow end. In other cases, the particles appear to join end to end. This latter situation usually happened with the thicker ends of the U-shaped particle.

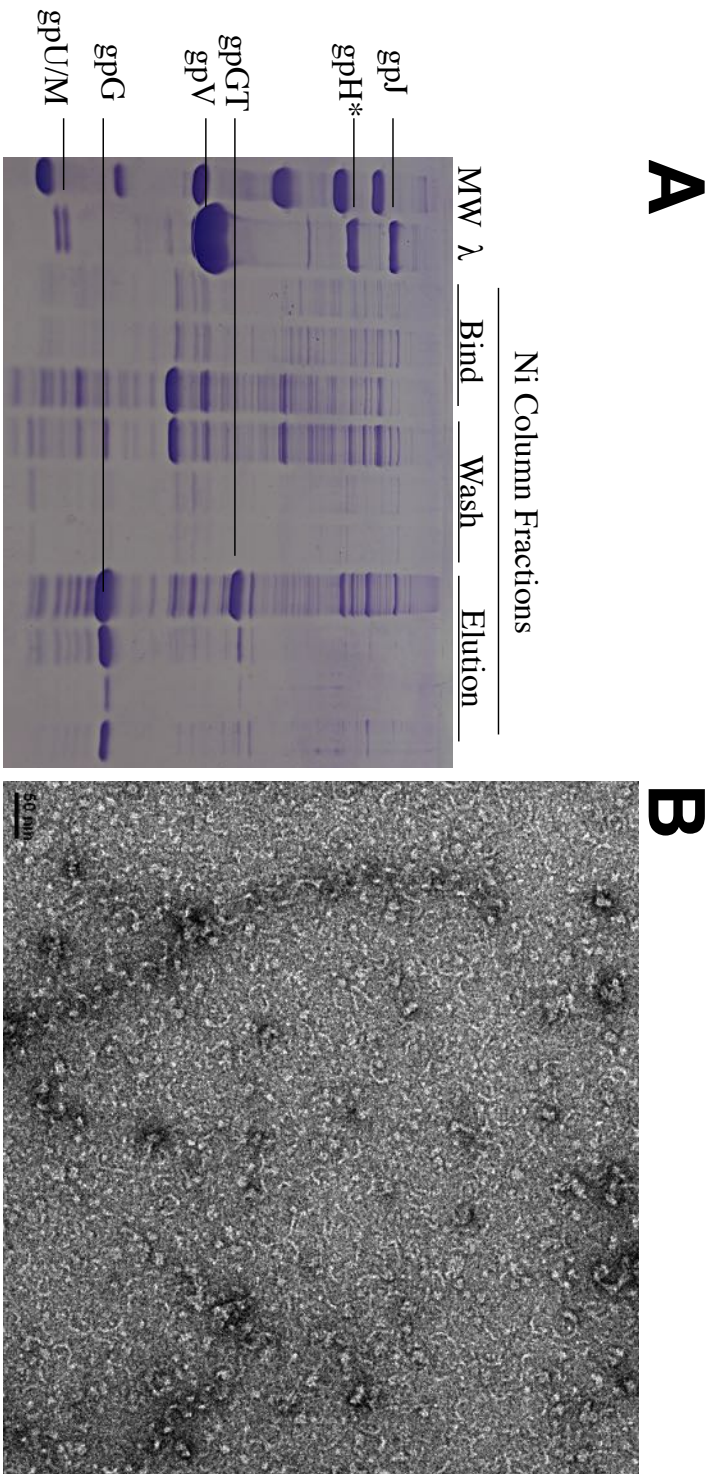

Figure 74. Purification and images of λ tail proteins in the initiator complex

A) 12.5% SDS showing nickel column fractions. MW is a molecular weight ladder. (λ) is the purified tail fraction. Bands with similar mobilities as tail proteins are shown in the elution fractions. B) Negative stain TEM image of the nickel column elution fraction in A. We can see a repeating looping structure along with individual loop structures in the background.

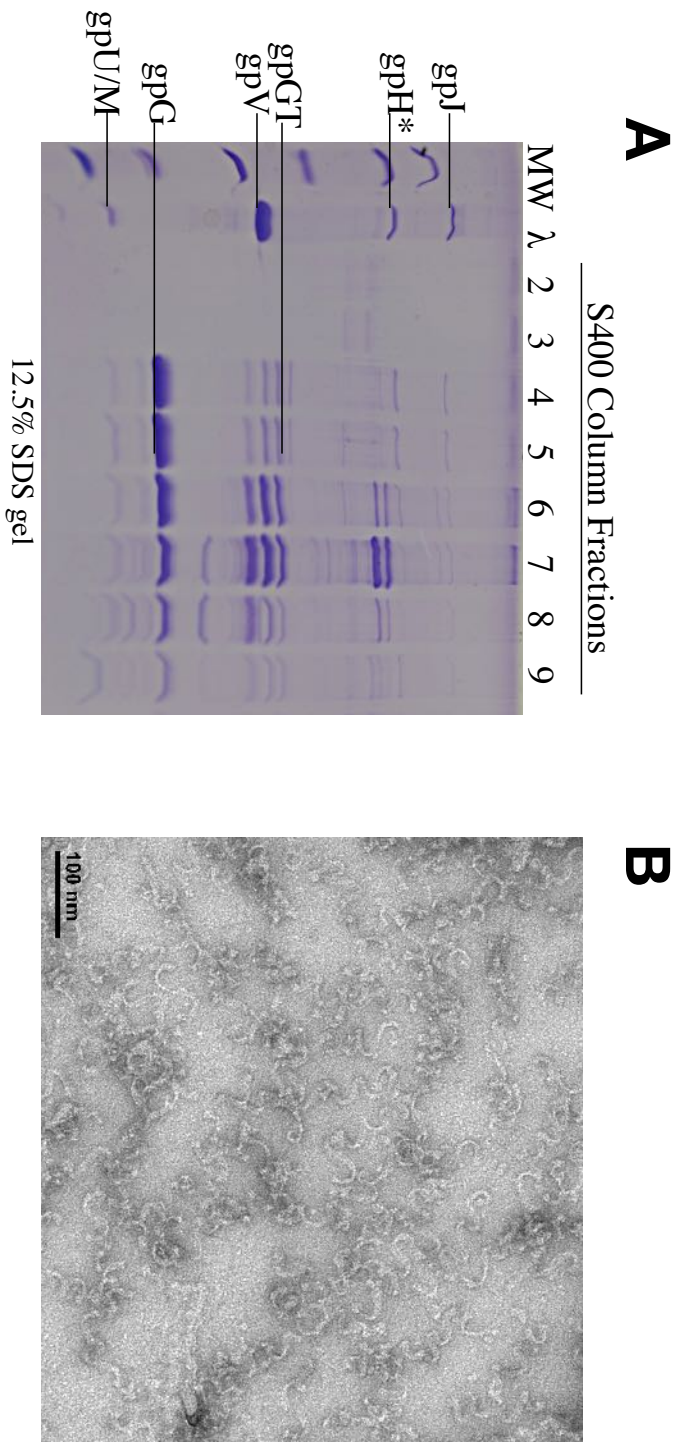

Figure 75. Size exclusion chromatography of the putative tail assembly intermediate and EM images

A) 12.5% SDS showing S400 void volume fractions. (λ) is purified tail fraction. Bands with similar mobilities as tail proteins are shown in fraction 4. Additional bands in other fractions were identified as tail proteins using mass spectrometry. B) Negative stain TEM image of void volume fraction 4. Some of the loop structures are connected while individual U-shape structures are present.

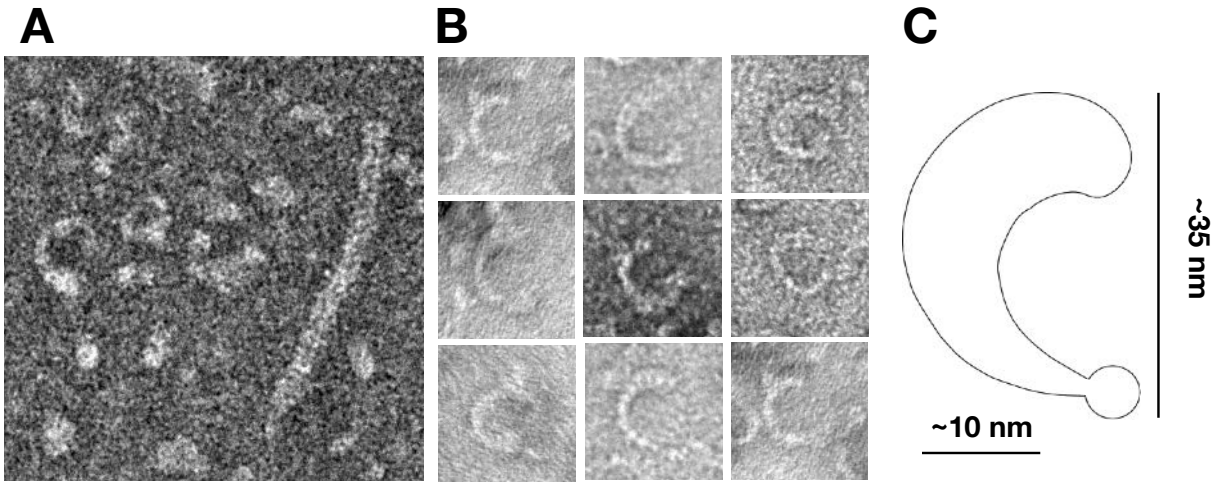

Figure 76. The basic structure of the putative tail assembly intermediate

A) An EM image of a Ni column fraction mixed with a purified tail sample for direct size comparison is shown. B) A gallery of images of the complexes observed in the fourth S400 size exclusion column fraction is shown. C) A simple representation of the complex with lengths shown.

6.2.2 The fates of λ tail proteins during assembly as assessed by mass spectrometry

Previous studies suggested that the TTC proteins gpJ and gpL are both present in the conical tail tip, whereas the presence and stoichiometry of gpK and gpI are uncertain (Casjens & Hendrix, 1974; Dai, 2009; Roessner & Ihler, 1984; Tam *et al.*, 2013). SDS gel analysis suggests that gpK is not present in purified tails or the virion; and that gpI is only associated with purified tails (Dai, 2009). Previous studies also showed that the cleaved form of tmp, gpH*, is present in the virion tail tube. However, the location of the ~100 amino acid fragment that is cleaved from the C-terminus of gpH remained unknown (Hendrix & Casjens, 1974; Murialdo & Siminovitch, 1972; Tsui & Hendrix, 1983).

Mass spectrometry, performed at the University of Pittsburgh Biomedical Mass Spectrometry Center (BioMS), was used to determine whether each of λ's minor tail proteins form

a part of the virion structure or are transiently required during assembly and discarded. The protein composition of the putative assembly intermediate, the purified tail and the virion submitted for mass spectrometry analysis. Samples were made by running the proteins present in SDS gel samples only a few mm into the gel, excising all proteins present together as a tight cluster of bands and submitting them form mass analysis. For the "assembly complex" analysis we used S400 void volume fractions 4, 6, 7 and 9 (Figure 75A) because they appeared representative of the proteins observed. Samples of purified λ tail, whole λ virion, as well as a gpH* band were also submitted for analysis.

Table 2. Copy numbers for λ tail proteins as estimated by densitometry (Dai 2009)

	λ tail	λ phage
gpJ	3	3
gpI	1-3	
gpL	3-4	~1
gpK		
gpH	1-3	
gpH*	5-8	6
gpV	153-177	178
gpU		6
gpM		8

6.2.2.1 gpK is lost early in the assembly pathway

The TTC protein gpK was not detected by mass spectrometry in the putative assembly intermediate sample, purified tails or virions (Table 3). This is in line with SDS gel analysis that did not detect gpK in either the purified tails or virions (Table 2)(Dai, 2009).

6.2.2.2 The C-terminal 100 residues of gpH were not detected in the virion

Analysis of the gpH* present in the purified tail sample showed that the C-terminal 100 residues were missing confirming that the conversion of gpH to gpH* during assembly is due to the loss of the C-terminus (Murialdo & Siminovitch, 1972)(Figure 77). When the protein composition of the putative tail assembly intermediate and purified tail sample was analyzed, the C-terminal 100 residues were present. However, the C-terminal 100 residues were not found in the virion sample.

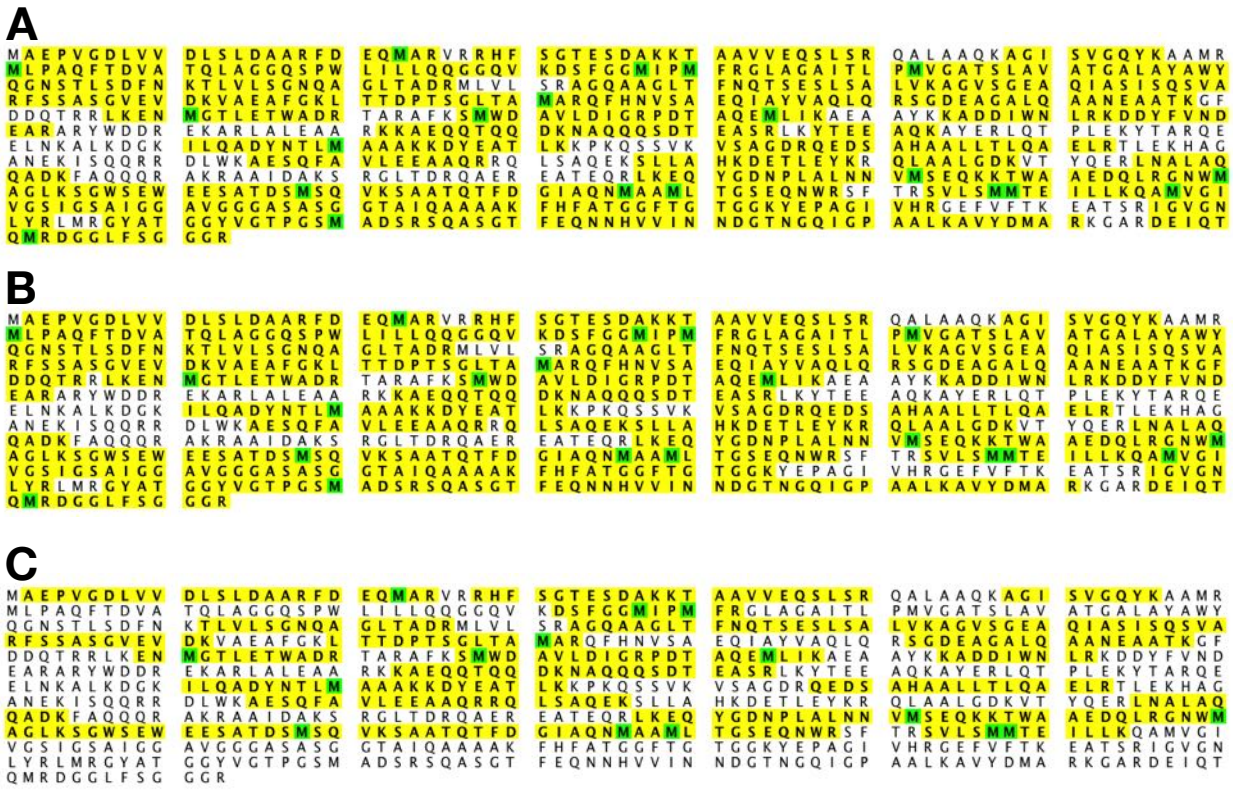

Figure 77. Mass-spectrometry coverage of gpH in samples of the assembly complex, purified tails, and virions

Areas of the amino acid sequence that are in yellow are sections that were detected in the mass spectrometry experiment. Images were obtained from the Scaffold software suite (Searle, 2010). A) The coverage of gpH present in the fourth fraction of a S400 size exclusion column. B) The coverage of gpH present in a purified tail sample. C) The coverage of gpH present in virions.

6.2.2.3 gpI is not detected after Head attachment

The TTC protein gpI is present in the putative tail assembly intermediate and the purified tail but not the virion (Table 3,4). The loss of gpI after head attachment was also observed in previous work (Dai, 2009).

6.2.2.4 The putative tail assembly intermediates are comprised of tail proteins

The S400 void volume samples are comprised of the tail proteins, gpG, gpGT, gpH, gpM, gpJ, gpI, and gpL which are proteins expected to be present in the initiator complex. The mass

spectrometry results strongly suggest that additional SDS protein bands present in the S400 column samples appear to be cleaved forms of gpH, gpG, and gpGT, suggesting that proteases were present during the purification process.

6.2.2.5 Other results

Surprisingly, gpU and gpZ were not detected in the virion sample (Table 3,4) although they are both expected to be present. In addition, the tmp chaperones gpG and gpGT were present in the purified tail sample, suggesting that the tail sample was impure. The protein that comprises the side tail fibers, gp*stf*, was detected in the virion sample but only the N-terminal region that has been defined as gpORF 401 (Figure 79)(Haggård-Ljungquist *et al.*, 1992).

Figure 78. Mass spectrometry detection of ORF 401 in the virion TTC

Areas of the amino acid sequence that are in yellow are sections that were detected in the mass spectrometry experiment. Images were obtained from the Scaffold software suite (Searle, 2010).

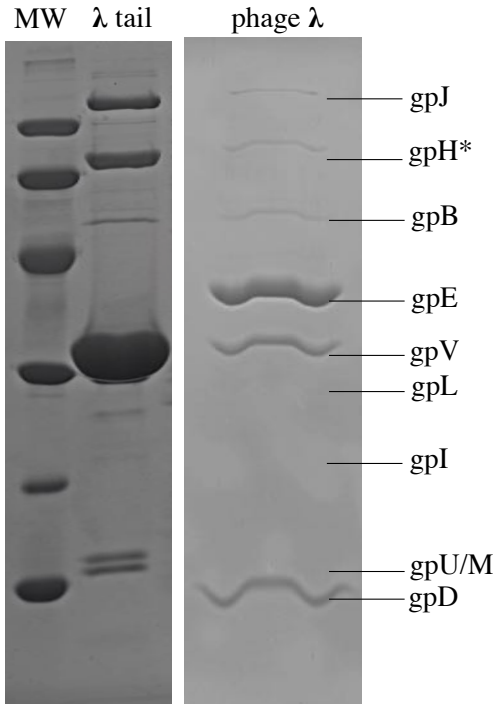

Figure 79. SDS gels of samples used for mass spectrometry

Table 3. Percent coverage of λ proteins present in assembly complex, purified tail, and virion samples

			Assembly complex (S400 samples)				whole Tail samples		Mature Virion
		MS-IDs	RHX-F4	RHX-F6	RHX-F7	RHX-F9	RHX-45	RHX-28	Law_RHX
	Protein	Accession ID							
Tail proteins	Z	C6ZCY5					38	31	
	U	C6ZCY6					42	27	
	H	C6ZCZ0	78	68	45	50	78	77	46
	V	C6ZCY7					89	76	75
	GT	P03735	80	80	80	80	65	78	
	M	C6ZCZ1	75	39	31	42	80	65	75
	J	C6ZCZ5	20	35	7.3	36	69	65	37
	K								
	I	C6ZCZ4	46	40	30	34	41	46	
	L	C6ZCZ2	27	36	6	60	88	75	62
	STF (ORF 401)	P03764							37
	tfa								
Head and Head-Joining proteins	E	C6EJ25							89
	B	C6EJF1							68
	D	C6EJE9							90
	C	C6EJF0							15
	Fi	C6EJ24							54
	Fii	C6EJ23							39
	W	C6EJF2							21

Table 4. Summary of proteins in λ tail samples

Sample	Tube	Tail tip complex
Assembly Intermediate	gpG, gpGT, gpH	gpJ, gpI, gpL, gpM
Purified tail	gpGT, gpH, gpZ, gpU, gpV	gpJ, gpI, gpL, gpM
Virion	gpV, gpH missing the last 100 residues, Fii	gpJ, gpL, gpM, gp*stf* (gpORF 401)

6.3 Discussion

The putative tail assembly intermediate contains gpH and its chaperones along with several TTC proteins. The distinct U-shaped particles are no longer than 50 nm tip to tip, suggesting a compressed state of the tmp. The sedimentation coefficients of the assembly intermediates (Katsura, 1976b; Katsura & Kuhl, 1975) suggest that the tmp is in a compact form in the initiator complex and extended during gpV polymerization. The putative assembly intermediates we observed are smaller and more compact than purified tails, which agrees with that prediction. When the D61A/D62A gpV mutant is coexpressed with WT λ phage, tail assembly is prematurely halted during polymerization (Pell, 2010; Pell *et al.*, 2009). No full-length structures are observed in EM images of these partially assembled tails, further indicating that the tmp is not in a fully extended form prior to gpV polymerization or during gpV polymerization (Pell, 2010; Pell *et al.*, 2010). What is not apparent in the complexes detailed here is the distinct conical tail tip. This may be due to structural rearrangements of the TTC proteins during assembly or more likely that the tail-tip proteins are loosely associated with the complex and dissociate easily. If the TTC proteins are only loosely associated with the complex, the shapes we observe may contain only the tmp and its chaperones. Clearly, further work needs to be done to confirm that this complex is biologically relevant. Such work would need to proceed carefully as the pTQ30 plasmid is a system that overexpresses proteins and tail assembly is "extremely concentration dependent" (Kuhl & Katsura, 1975).

The TTC of the λ tail may undergo structural changes during head-tail attachment. The TTC proteins gpK and gpI are required for assembly, but their location and function is still unknown. When all the TTC proteins are expressed together, of the proteins gpI, gpJ, and gpL appear to make a stable complex but it does not appear that gpK is a part of that complex (Dai,

2009). In this work, the complex we purified included many of the proteins expected in the initiator complex, but gpK was not detected. We also did not detect gpK in purified tails or virions. The TTC protein gpI was detected in the putative tail-assembly intermediates and in purified tails, but not in virions. The present of gpG and gpGT in the purified tail sample suggests that the presence of gpI may be due to the sample being impure. However, this loss of gpI after head attachment was observed before (Dai, 2009). The explanation for the difference in protein composition between purified tails and virion tails is still not clear.

Contrary to expectations, proteins gpU and gpZ were detected in the purified tail samples, but not in the virion. While gpZ was not observed in virions, gpU was observed in the virion. In previous work (Dai, 2009) gpU was detected as a gel band in the virion, but gpZ was not. It's not clear why gpU was not detected in the virion samples here. The presence of gpU in the virion has often been assumed (Edmonds *et al.*, 2007; Tsui & Hendrix, 1980), so it is possible that the lack of detection may be due to a technical issue, such the loss of gpU during the preparation of gel bands for mass spectrometry. The presence of gp*stf* was surprising because the 'wild type' phage λ does not have side tail fibers (Hendrix & Duda, 1992) due to a frameshift mutation in the *stf* gene. The frameshifted *stf* gene is predicted to make 2 proteins, gpORF 314 and gpORF 401. Mass spectrometry analysis showed considerable coverage of the N-terminal region of gp*stf* corresponding almost exactly to ORF 401 (Figure 78). This shows that gpORF 401 is capable of binding to the λ-PaPa TTC in an appreciable quantity. This may explain the 'dots' that were observed in cryo-EM images of the virion, as described in Chapter 5.

This work suggests that the tmp C-terminus is lost during head attachment. The cleavage of the tmp occurs during the termination of gpV polymerization (Murialdo & Siminovitch, 1972; Tsui & Hendrix, 1983). While the cleavage of the tmp during assembly has long been accepted,

the purpose of the tmp cleavage as well as the fate of the cleavage product is not known. This work suggests that the C-terminus, while cleaved from gpH is still present in some capacity in the purified tail sample but is lost after head attachment. The presence of the tmp chaperones counters this observation as those were not expected to be present after gpV polymerization. Once again, it is possible that the c-terminus is actually present in the virion but was simply not detected. If it is lost after head attachment, this confines the role of the λ tmp C-terminus to assembly only since it is required for the formation of the initiator (Katsura, 1990; Murialdo & Siminovitch, 1972).

6.4 Future directions

The biological relevance of the putative tail assembly intermediate isolated in this study needs to be determined. Since tail assembly is highly sensitive to the concentration of the tail proteins, in-vitro complementation may be the ideal way to test the activity of this complex because, unlike in-vivo complementation, the relative concentrations of proteins can be better controlled. Another way to test the activity of the complex is to see if polytails are formed when combined with monomeric gpV. Since monomeric gpV can be purified (Pell *et al.*, 2009; Xu *et al.*, 2014), such a study could be performed in the future. If it is possible to obtain the initiator complex, a structural study of the polymerization of gpV onto the initiator can be performed using 2D averaging. One way to achieve this would be to mix a sample of the initiator complex with a sample of monomeric gpV, allowing a short period of time to pass, applying the sample to a cryo-grid then plunge-freezing the sample. Ideally, this will allow visualization of gpV polymerizing on the initiator which can be used to generate a 2D average. This approach may reveal how the chaperones, tmp and gpV interact.

The role of gpK is still a mystery. In K- lysates, there may be "K- particles" with the same sedimentation velocity of regular λ phage particles (Kuhl & Katsura, 1975). gpK may be a part of the CHAP (cysteine, histidine-dependent amidohydrolases/peptidases) superfamily (Rigden *et al.*, 2003) so it could potentially be a candidate for gpH proteolysis. Testing whether gpK has this enzymatic activity and determining whether gpH is cleaved in the absence of gpK, could provide insight into its role in assembly.

7.0 Discussion

In this work, I provided initial evidence that a virus sharing ~50% sequence identity with the phage HK97 mcp employs different networks of contacts to generate the larger phi1026b capsid. While the significance of individual differences in the contact network remains to be established, the clear conclusion is that no single change determines the difference in capsid geometry. I also showed that 'skewing' appears to be a conserved element of capsomers in proheads, but that the details of skewing adapt to the local symmetry of the capsomer location. A third major outcome of my work is that the delta domains also adopt recognizably similar assembly conformations while also obeying the constraints of three different positional symmetries – icosahedral 5-fold and 3-fold, and local 2-fold. Despite the high sequence similarity between phages HK97 and phi1026b in the mcps, there is structural variability as well as flexibility in local interaction that occur between identical copies of the same protein such that a larger capsid is assembled. While the work presented here increases understanding of phage capsid structure and assembly, future work may be expected to fill in gaps between sequence and structure by revealing how the balance between variability and conservation yields precision assembly of unique capsid geometries.

At the completion of this thesis, only one other high-resolution structure for a *Siphoviridae* tail tube was available, that of phage T5 at 6Å resolution (Arnaud *et al.*, 2017). I was fortunate to have atomic models of the λ gpV major tail protein domains available that allowed construction of very detailed models of the polymerized tail tube. The most exciting result was showing that the λ tail tube assembly and structure is guided by electrostatics involving not only the tail tube N-terminal domain but also the C-terminal Ig-like domains. Additionally, I was able to present here

a moderately high resolution model of a *Siphoviridae* tail tip complex (TTC) that is validated by visualizing density from the end f the tail tube that is consistent with the gpV fold and arrangement of the Ig-like domains observed in the tail tube structure. I observed that the λ TTC obeys 3-fold symmetry, connects to the 6-fold symmetric tail tube, and is structurally similar to a model of the T4 inner baseplate suggesting an evolutionary relationship in the organization of tail tips in *Siphoviridae* and *Myoviridae*. Finally, this work also presents a stable, putative *Siphoviridae* tail assembly intermediate.

Overall, this work adds to the knowledge of how proteins coordinate to generate large macromolecular structures. The capsid studies reveal a protein fold that adapts to different icosahedral geometries (T=7 and T=9), while the individual proteins are able to adapt to pentameric organization, as well as hexameric exhibiting pseudo-2-fold symmetry and hexameric obeying strict 3-fold symmetry. Presumably it is this ability to associate in multiple ways that allows for a conserved protein fold to evolve from assembling smaller capsids to assembling successively larger capsids. The work on the λ tail tube also details a conserved protein fold that is found both in phage tail tubes as well as in bacterial systems, and which assembles in structure of varying rigidity such as the tail tubes of λ that is flexible and T4 that is rigid. This varying large-scale rigidity appears to result from differences in the degree of contact between successive layers of tail-tube disks in the assembled tube and shows how a single protein fold can be adapted to generate structures with alternate physical properties.

I carried out my thesis work during a time of exponential growth in the power of cryo-EM techniques – almost every year, a new and improved data processing and/or reconstruction software package was made available allowing significant jumps in the ability to answer research questions. While this rapid progress was very exciting, it also presented a challenge as to when to

put a pin on one's research. On one hand, a breakthrough one year may be considered routine the next while on the other hand, revisiting older but good quality data might answer questions that were previously out of reach, especially those affected by structural flexibility and uniformity. What is presented is thus a compromise and may be quickly surpassed by further success in the field.

My hope for this work going forward is for additional experiments that will substantiate and extend the conclusions made here. In order to better understand how the highly similar phi1026b mcp featuring the HK97 mcp fold is adapted to a larger capsid, published experiments conducted on the HK97 mcp should be repeated with phi1026b, focusing particularly on amino acids that affect assembly. The goal would be to more fully define the networks of interactions and how the changes to amino acids work together to specify different capsid geometries. The structure of the phage λ tail tube without the Ig-like domains would be especially interesting to determine – for example, would their absence affect the helical parameters of the tail tube? For extending the λ TTC work, a significant hurdle is collecting high-quality datasets with a sufficient number of particles to satisfy classifications according to variable conformations. However, if this could be accomplished, then a detailed model of this host-recognition and penetration complex could be determined that would offer insights into the mechanism of phage infection. Finally, it is my hope that the tail assembly intermediate will be confirmed and better characterized as a step for understanding the complex process of tail assembly. This process is also relevant for studying of bacterial systems that are important for human health, such as the secretion systems in bacteria.

Bibliography

Abuladze, N. K., Gingery, M., Tsai, J., & Eiserling, F. A. (1994). Tail Length Determination in Bacteriophage T4. *Virology, 199*(2), 301-310.

Ackermann, H. W. (2011). The first phage electron micrographs. *Bacteriophage, 1*(4), 225-227.

Ackermann, H. W., & Prangishvili, D. (2012). Prokaryote viruses studied by electron microscopy. *Archives of Virology, 157*(10), 1843-1849.

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 2), 213-221.

Adrian, M., Dubochet, J., Fuller, S. D., & Harris, J. R. (1998). Cryo-negative staining. *Micron, 29*(2), 145-160.

Akita, F., Chong, K. T., Tanaka, H., Yamashita, E., Miyazaki, N., Nakaishi, Y., . . . Nakagawa, A. (2007). The Crystal Structure of a Virus-like Particle from the Hyperthermophilic Archaeon Pyrococcus furiosus Provides Insight into the Evolution of Viruses. *Journal of Molecular Biology, 368*(5), 1469-1483.

Aksyuk, A. A., & Rossmann, M. G. (2011). Bacteriophage assembly. *Viruses, 3*(3), 172-203.

Arnaud, C.-A., Effantin, G., Vivès, C., Engilberge, S., Bacia, M., Boulanger, P., . . . Breyton, C. (2017). Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection. *Nature communications, 8*(1), 1953-1953.

Bachrach, U., & Friedmann, A. (1971). Practical procedures for the purification of bacterial viruses. *Applied microbiology, 22*(4), 706-715.

Baker, M. L., Jiang, W., Rixon, F. J., & Chiu, W. (2005). Common ancestry of herpesviruses and tailed DNA bacteriophages. *Journal of virology, 79*(23), 14967-14970.

Baker, T. S., Olson, N. H., & Fuller, S. D. (1999). Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiology and molecular biology reviews : MMBR, 63*(4), 862-922.

Bammes, B. E., Rochat, R. H., Jakana, J., & Chiu, W. (2011). Practical performance evaluation of a 10k×10k CCD for electron cryo-microscopy. *Journal of Structural Biology, 175*(3), 384-393.

Baschong, W., Aebi, U., Baschong-Prescianotto, C., Dubochet, J., Landmann, L., Kellenberger, E., & Wurtz, M. (1988). Head structure of bacteriophages T2 and T4. *Journal of Ultrastructure and Molecular Structure Research, 99*(3), 189-202.

Bebeacua, C., Bron, P., Lai, L., Vegge, C. S., Brøndsted, L., Spinelli, S., . . . Cambillau, C. (2010). Structure and molecular assignment of lactococcal phage TP901-1 baseplate. *The Journal of biological chemistry, 285*(50), 39079-39086.

Bell, J. M., Chen, M., Baldwin, P. R., & Ludtke, S. J. (2016). High resolution single particle refinement in EMAN2.1. *Methods (San Diego, Calif.), 100*, 25-34.

Bergh, Ø., BØrsheim, K. Y., Bratbak, G., & Heldal, M. (1989). High abundance of viruses found in aquatic environments. *Nature, 340*(6233), 467-468.

Bleviss, M., & Easterbrook, K. B. (1971). Self-assembly of bacteriophage lambda tails. *Canadian Journal of Microbiology, 17*(7), 947-954.

Boulanger, P., Jacquot, P., Plancon, L., Chami, M., Engel, A., Parquet, C., . . . Letellier, L. (2008). Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. *J Biol Chem, 283*(20), 13556-13564.

Bradley, D. E. (1962). A Study of the Negative Staining Process. *Journal of General Microbiology, 29*(3), 503-516.

Bradley, D. E. (1967). Ultrastructure of bacteriophage and bacteriocins. *Bacteriological reviews, 31*(4), 230-314.

Brussaard, C. P. D., Wilhelm, S. W., Thingstad, F., Weinbauer, M. G., Bratbak, G., Heldal, M., . . . Wommack, K. E. (2008). Global-scale processes with a nanoscale drive: the role of marine viruses. *The ISME Journal, 2*(6), 575-578.

Brüssow, H., & Hendrix, R. W. (2002). Phage Genomics: Small Is Beautiful. In (Vol. 108, pp. 13-16). United States: Elsevier Inc.

Carragher, B., Kisseberth, N., Kriegman, D., Milligan, R. A., Potter, C. S., Pulokas, J., & Reilein, A. (2000). Leginon: An Automated System for Acquisition of Images from Vitreous Ice Specimens. *Journal of Structural Biology, 132*(1), 33-45.

Casjens, S., & King, J. (1975). Virus Assembly. *Annual Review of Biochemistry, 44*(1), 555-611.

Casjens, S. R., & Hendrix, R. W. (1974). Locations and amounts of the major structural proteins in bacteriophage lambda. *Journal of Molecular Biology, 88*(2), 535-545.

Casjens, S. R., & Hendrix, R. W. (1988). Control mechanisms in dsDNA bacteriophage assembly. In R. Calendar (Ed.), *The Bacteriophages* (Vol. 1, pp. 15-91). New York City: Plenum Press.

Casjens, S. R., & Molineux, I. J. (2012). Short Noncontractile Tail Machines: Adsorption and DNA Delivery by Podoviruses. In (2012 ed., Vol. 726, pp. 143-179). Boston, MA: Springer US.

Caspar, D. L., & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol, 27*, 1-24.

Chang, J., Liu, X., Rochat, R. H., Baker, M. L., & Chiu, W. (2012). Reconstructing virus structures from nanometer to near-atomic resolutions with cryo-electron microscopy and tomography. *Advances in experimental medicine and biology, 726*, 49-90.

Chen, Z., Sun, L., Zhang, Z., Fokine, A., Padilla-Sanchez, V., Hanein, D., . . . Rao, V. B. (2017). Cryo-EM structure of the bacteriophage T4 isometric head at 3.3-Å resolution and its relevance to the assembly of icosahedral viruses. *Proceedings of the National Academy of Sciences of the United States of America, 114*(39), E8184-E8193.

Cho, H.-J., Hyun, J.-K., Kim, J.-G., Jeong, H. S., Park, H. N., You, D.-J., & Jung, H. S. (2013). Measurement of ice thickness on vitreous ice embedded cryo-EM grids: investigation of optimizing condition for visualizing macromolecules. *Journal of Analytical Science and Technology, 4*(1), 7.

Christie, G. E., & Dokland, T. (2012). Pirates of the Caudovirales. *Virology, 434*(2), 210-221.

Conway, J. F., Cheng, N., Ross, P. D., Hendrix, R. W., Duda, R. L., & Steven, A. C. (2007). A thermally induced phase transition in a viral capsid transforms the hexamers, leaving the pentamers unchanged. *Journal of Structural Biology, 158*(2), 224-232.

Conway, J. F., Duda, R. L., Cheng, N., Hendrix, R. W., & Steven, A. C. (1995). Proteolytic and Conformational Control of Virus Capsid Maturation: The Bacteriophage HK97 System. *Journal of Molecular Biology, 253*(1), 86-99.

Conway, J. F., & Steven, A. C. (1999). Methods for Reconstructing Density Maps of "Single" Particles from Cryoelectron Micrographs to Subnanometer Resolution. *Journal of Structural Biology, 128*(1), 106-118.

D'Herelle, F. (1917). Sur un microbe invisible antagoniste des bacilles dysentériques. *C. R. Academy of Sciences, Paris, 165*, 373-375.

Dai, X. (2009). *Expression, purification and characterization of bacteriophage lambda tail tip proteins.* (Ph.D. Ph.D.), University of Pittsburgh,

Davidson, A. R., Cardarelli, L., Pell, L. G., Radford, D. R., & Maxwell, K. L. (2012). Long Noncontractile Tail Machines of Bacteriophages. In (2012 ed., Vol. 726, pp. 115-142). Boston, MA: Springer US.

De Carlo, S., Boisset, N., & Hoenger, A. (2008). High-resolution single-particle 3D analysis on GroEL prepared by cryo-negative staining. *Micron, 39*(7), 934-943.

De Carlo, S., El-Bez, C., Alvarez-Rua, C., Borge, J., & Dubochet, J. (2002). Cryo-negative staining reduces electron-beam sensitivity of vitrified biological particles. *J Struct Biol, 138*(3), 216-226.

De Carlo, S., & Harris, J. R. (2011). Negative staining and cryo-negative staining of macromolecules and viruses for TEM. *Micron, 42*(2), 117-131.

Desfosses, A., Ciuffa, R., Gutsche, I., & Sachse, C. (2014). SPRING - an image processing package for single-particle based helical reconstruction from electron cryomicrographs. *J Struct Biol, 185*(1), 15-26.

DeShazer, D. (2004). Genomic diversity of Burkholderia pseudomallei clinical isolates: subtractive hybridization reveals a Burkholderia mallei-specific prophage in B. pseudomallei 1026b. *Journal of bacteriology, 186*(12), 3938-3950.

Dhillon, E. K., Dhillon, T. S., Lai, A. N., & Linn, S. (1980). Host range, immunity and antigenic properties of lambdoid coliphage HK97. *J Gen Virol, 50*(1), 217-220.

Dierkes, L. E., Peebles, C. L., Firek, B. A., Hendrix, R. W., & Duda, R. L. (2009). Mutational Analysis of a Conserved Glutamic Acid Required for Self-Catalyzed Cross-Linking of Bacteriophage HK97 Capsids. *Journal of virology, 83*(5), 2088.

Ding, Y., Duda, R. L., Hendrix, R. W., & Rosenberg, J. M. (1995). Complexes between chaperonin GroEL and the capsid protein of bacteriophage HK97. *Biochemistry, 34*(45), 14918-14931.

Dokland, T. (1999). Scaffolding proteins and their role in viral assembly. *Cellular and Molecular Life Sciences, 56*, 580-603.

Dokland, T., & Murialdo, H. (1993). Structural Transitions During Maturation of Bacteriophage Lambda Capsids. *Journal of Molecular Biology, 233*(4), 682-694.

Dowah, A. S. A., & Clokie, M. R. J. (2018). Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria. *Biophysical reviews, 10*(2), 535-542.

Dubochet, J., Adrian, M., Lepault, J., & McDowall, A. W. (1985). Emerging techniques: Cryo-electron microscopy of vitrified biological specimens. *Trends in Biochemical Sciences, 10*(4), 143-146.

Dubochet, J., Lepault, J., Freeman, R., Berriman, J. A., & Homo, J. C. (1982). Electron microscopy of frozen water and aqueous solutions. *Journal of Microscopy, 128*(3), 219-237.

Duda, R. L. (1998). Protein chainmail: catenated protein in viral capsids. *Cell, 94*(1), 55-60.

Duda, R. L., Hempel, J., Michel, H., Shabanowitz, J., Hunt, D., & Hendrix, R. W. (1995). Structural transitions during bacteriophage HK97 head assembly. *J Mol Biol, 247*(4), 618-635.

Duda, R. L., Martincic, K., & Hendrix, R. W. (1995). Genetic basis of bacteriophage HK97 prohead assembly. *Journal of Molecular Biology, 247*(4), 636-647.

Duda, R. L., Martincic, K., Xie, Z., & Hendrix, R. W. (1995). Bacteriophage HK97 head assembly. *FEMS Microbiol Rev, 17*(1-2), 41-46.

Duda, R. L., Oh, B., & Hendrix, R. W. (2013). Functional domains of the HK97 capsid maturation protease and the mechanisms of protein encapsidation. *Journal of Molecular Biology, 425*(15), 2765-2781.

Edgar, R. S., & Lielausis, I. (1968). Some steps in the assembly of bacteriophage T4. *Journal of Molecular Biology, 32*(2), 263-276.

Edmonds, L., Liu, A., Kwan, J. J., Avanessy, A., Caracoglia, M., Yang, I., . . . Donaldson, L. W. (2007). The NMR Structure of the gpU Tail-terminator Protein from Bacteriophage Lambda: Identification of Sites Contributing to Mg(II)-mediated Oligomerization and Biological Function. *Journal of Molecular Biology, 365*(1), 175-186.

Egelman, E. H. (2007). The iterative helical real space reconstruction method: surmounting the problems posed by real polymers. *J Struct Biol, 157*(1), 83-94.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 4), 486-501.

Flayhan, A., Vellieux, F. M. D., Lurz, R., Maury, O., Contreras-Martel, C., Girard, E., . . . Breyton, C. (2014). Crystal structure of pb9, the distal tail protein of bacteriophage T5: a conserved structural motif among all siphophages. *Journal of virology, 88*(2), 820-828.

Frank, J. (2017). Advances in the field of single-particle cryo-electron microscopy over the last decade. *Nature Protocols, 12*, 209.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., & Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol, 116*(1), 190-199.

Fraser, J. S., Yu, Z., Maxwell, K. L., & Davidson, A. R. (2006). Ig-like domains on bacteriophages: a tale of promiscuity and deceit. *J Mol Biol, 359*(2), 496-507.

Gan, L., Speir, J. A., Conway, J. F., Lander, G., Cheng, N., Firek, B. A., . . . Johnson, J. E. (2006). Capsid conformational sampling in HK97 maturation visualized by X-ray crystallography and cryo-EM. *Structure, 14*(11), 1655-1665.

Gertsman, I., Gan, L., Guttman, M., Lee, K., Speir, J. A., Duda, R. L., . . . Johnson, J. E. (2009). An unexpected twist in viral capsid maturation. *Nature, 458*(7238), 646-650.

Gong, Y., Veesler, D., Doerschuk, P. C., & Johnson, J. E. (2016). Effect of the viral protease on the dynamics of bacteriophage HK97 maturation intermediates characterized by variance analysis of cryo EM particle ensembles. *Journal of Structural Biology, 193*(3), 188-195.

Goulet, A., Lai-Kee-Him, J., Veesler, D., Auzat, I., Robin, G., Shepherd, D. A., . . . Bron, P. (2011). The opening of the SPP1 bacteriophage tail, a prevalent mechanism in Gram-positive-infecting siphophages. *The Journal of biological chemistry, 286*(28), 25397-25405.

Grose, J. H., & Casjens, S. R. (2014). Understanding the enormous diversity of bacteriophages: The tailed phages that infect the bacterial family Enterobacteriaceae. *Virology, 468*, 421-443.

Guerrini, N., Turchetta, R., Hoften, G. V., Henderson, R., McMullan, G., & Faruqi, A. R. (2011). A high frame rate, 16 million pixels, radiation hard CMOS sensor. *Journal of Instrumentation, 6*(3), C03003-C03003.

Guex, N., & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis, 18*(15), 2714-2723.

Habenstein, B., & Loquet, A. (2016). Solid-state NMR: An emerging technique in structural biology of self-assemblies. *Biophysical Chemistry, 210*, 14-26.

Habenstein, B., Loquet, A., Giller, K., Becker, S., & Lange, A. (2013). Structural characterization of supramolecular assemblies by (1)(3)C spin dilution and 3D solid-state NMR. *J Biomol NMR, 55*(1), 1-9.

Haggård-Ljungquist, E., Halling, C., & Calendar, R. (1992). DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. *Journal of bacteriology, 174*(5), 1462-1477.

Hasek, M. L., Maurer, J. B., Hendrix, R. W., & Duda, R. L. (2017). Flexible Connectors between Capsomer Subunits that Regulate Capsid Assembly. *Journal of Molecular Biology, 429*(16), 2474-2489.

Helgstrand, C., Wikoff, W. R., Duda, R. L., Hendrix, R. W., Johnson, J. E., & Liljas, L. (2003). The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 A resolution. *J Mol Biol, 334*(5), 885-899.

Hendrix, R. W. (2003). Bacteriophage genomics. *Current Opinion in Microbiology, 6*(5), 506-511.

Hendrix, R. W., & Casjens, S. R. (1974). Protein cleavage in bacteriophage λ tail assembly. *Virology, 61*(1), 156-159.

Hendrix, R. W., & Duda, R. L. (1992). Bacteriophage lambda-PaPa: not the mother of all lambda phages. *Science, 258*, 1145+.

Hendrix, R. W., & Johnson, J. E. (2012). Bacteriophage HK97 capsid assembly and maturation. *Adv Exp Med Biol, 726*, 351-363.

Heymann, J. B., & Belnap, D. M. (2007). Bsoft: Image processing and molecular modeling for electron microscopy. *Journal of Structural Biology, 157*(1), 3-18.

Hohn, T., & Katsura, I. (1977). *Structure and Assembly of Bacteriophage Lambda* (Vol. 78).

Homa, F. L., Huffman, J. B., Toropova, K., Lopez, H. R., Makhov, A. M., & Conway, J. F. (2013). Structure of the pseudorabies virus capsid: comparison with herpes simplex virus type 1 and differential binding of essential minor proteins. *Journal of Molecular Biology, 425*(18), 3415-3428.

Huang, R. K., Khayat, R., Lee, K. K., Gertsman, I., Duda, R. L., Hendrix, R. W., & Johnson, J. E. (2011). The Prohead-I structure of bacteriophage HK97: implications for scaffold-mediated control of particle assembly and maturation. *Journal of Molecular Biology, 408*(3), 541-554.

Huet, A., Duda, R. L., Hendrix, R. W., Boulanger, P., & Conway, J. F. (2016). Correct Assembly of the Bacteriophage T5 Procapsid Requires Both the Maturation Protease and the Portal Complex. *Journal of Molecular Biology, 428*(1), 165-181.

Huet, A., Makhov, A. M., Huffman, J. B., Vos, M., Homa, F. L., & Conway, J. F. (2016). Extensive subunit contacts underpin herpesvirus capsid stability and interior-to-exterior allostery. *Nature structural & molecular biology, 23*(6), 531-539.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph, 14*(1), 33-38, 27-38.

Kanamaru, S., Leiman, P. G., Kostyuchenko, V. A., Chipman, P. R., Mesyanzhinov, V. V., Arisaka, F., & Rossmann, M. G. (2002). Structure of the cell-puncturing device of bacteriophage T4. *Nature, 415*(6871), 553-557.

Kang, X., Kirui, A., Muszyński, A., Widanage, M. C. D., Chen, A., Azadi, P., . . . Wang, T. (2018). Molecular architecture of fungal cell walls revealed by solid-state NMR. *Nature communications, 9*(1), 2747-2747.

Katsura, I. (1976a). Isolation of λ prophage mutants defective in structural genes: Their use for the study of bacteriophage morphogenesis. *MGG Molecular & General Genetics, 148*(1), 31-42.

Katsura, I. (1976b). Morphogenesis of bacteriophage lambda tail. Polymorphism in the assembly of the major tail protein. *Journal of Molecular Biology, 107*(3), 307-326.

Katsura, I. (1981). Structure and function of the major tail protein of bacteriophage lambda: Mutants having small major tail protein molecules in their virion. *Journal of Molecular Biology, 146*(4), 493-512.

Katsura, I. (1983). Tail assembly and injection. In R. Hendrix, J. Roberts, F. Stahl, & R. Weisberg (Eds.), *Lambda II* (pp. 331-346). Cold Spring Harbor Laboratory, New York.

Katsura, I. (1990). Mechanism of length determination in bacteriophage lambda tails. *Advances in Biophysics, 26*, 1-18.

Katsura, I., & Hendrix, R. W. (1984). Length determination in bacteriophage lambda tails. *Cell, 39*(3), 691-698.

Katsura, I., & Kuhl, P. W. (1975). Morphogenesis of the tail of bacteriophage lambda. II. In vitro formation and properties of phage particles with extra long tails. *Virology, 63*(1), 238-251.

Katsura, I., & Kühl, P. W. (1975). Morphogenesis of the tail of bacteriophage lambda: III. Morphogenetic pathway. *Journal of Molecular Biology, 91*(3), 257-273.

Katsura, I., & Tsugita, A. (1977). Purification and characterization of the major protein and the terminator protein of the bacteriophage lambda tail. *Virology, 76*(1), 129-145.

Kellenberger, E., & Edgar, R. S. (1971). Structure and assembly phage particles. In A. D. Hershey (Ed.), *The Bacteriophage Lambda* (Vol. 2). Cold Spring Harbor Laboratory.

Keller, B., Dubochet, J., Adrian, M., Maeder, M., Wurtz, M., & Kellenberger, E. (1988). Length and shape variants of the bacteriophage T4 head: mutations in the scaffolding core genes 68 and 22. *Journal of virology, 62*(8), 2960-2969.

Kikuchi, Y., & King, J. (1975). Genetic control of bacteriophage T4 baseplate morphogenesis. I. Sequential assembly of the major precursor, in vivo and in vitro. *Journal of Molecular Biology, 99*(4), 645.

King, J. (1968). Assembly of the tail of bacteriophage T4. *J Mol Biol, 32*(2), 231-262.

Kleywegt, G. J., & Jones, T. A. (1996). xdlMAPMAN and xdlDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr D Biol Crystallogr, 52*(Pt 4), 826-828.

Koç, C., Xia, G., Kühner, P., Spinelli, S., Roussel, A., Cambillau, C., & Stehle, T. (2016). Structure of the host-recognition device of Staphylococcus aureus phage ϕ11. *Scientific reports, 6*, 27581-27581.

Kondou, Y., Kitazawa, D., Takeda, S., Tsuchiya, Y., Yamashita, E., Mizuguchi, M., . . . Tsukihara, T. (2005). Structure of the Central Hub of Bacteriophage Mu Baseplate Determined by X-ray Crystallography of gp44. *Journal of Molecular Biology, 352*(4), 976-985.

Konopa, G., & Taylor, K. (1979). Coliphage lambda ghosts obtained by osmotic shock or LiCl treatment are devoid of J- and H-gene products. *J Gen Virol, 43*(3), 729-733.

Kostyuchenko, V. A., Leiman, P. G., Chipman, P. R., Kanamaru, S., van Raaij, M. J., Arisaka, F., . . . Rossmann, M. G. (2003). Three-dimensional structure of bacteriophage T4 baseplate. *Nature Structural Biology, 10*, 688.

Krishnamurthy, S., Veesler, D., Khayat, R., Snijder, J., Huang, R., Heck, A., . . . Anand, G. S. (2014). Distinguishing direct binding interactions from allosteric effects in the protease-HK97 prohead I δ domain complex by amide H/D exchange mass spectrometry. *Bacteriophage, 4*(4), e959816-e959816.

Krupovič , M., & Bamford, D. H. (2008). Virus evolution: how far does the double β-barrel viral lineage extend? *Nature Reviews Microbiology, 6*(12), 941-948.

Kucukelbir, A., Sigworth, F. J., & Tagare, H. D. (2014). Quantifying the local resolution of cryo-EM density maps. *Nature methods, 11*(1), 63-65.

Kuhl, P. W., & Katsura, I. (1975). Morphogenesis of the tail of bacteriophage lambda I. In vitro intratail complementation. *Virology, 63*(1), 221-237.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature, 227*(5259), 680-685.

Lander, G. C., Baudoux, A.-C., Azam, F., Potter, C. S., Carragher, B., & Johnson, J. E. (2012). Capsomer dynamics and stabilization in the T = 12 marine bacteriophage SIO-2 and its procapsid studied by CryoEM. *Structure (London, England : 1993), 20*(3), 498-503.

Langlois, C., Ramboarina, S., Cukkemane, A., Auzat, I., Chagot, B., Gilquin, B., . . . Zinn-Justin, S. (2015). Bacteriophage SPP1 tail tube protein self-assembles into β-structure-rich tubes. *The Journal of biological chemistry, 290*(6), 3836-3849.

Lata, R., Conway, J. F., Cheng, N., Duda, R. L., Hendrix, R. W., Wikoff, W. R., . . . Steven, A. C. (2000). Maturation dynamics of a viral capsid: visualization of transitional intermediate states. *Cell, 100*(2), 253-263.

Lecoq, L., Wang, S., Wiegand, T., Bressanelli, S., Nassal, M., Meier, B. H., & Bockmann, A. (2018). Solid-state [(13)C-(15)N] NMR resonance assignment of hepatitis B virus core protein. *Biomol NMR Assign, 12*(1), 205-214.

Legrand, P., Collins, B., Blangy, S., Murphy, J., Spinelli, S., Gutierrez, C., . . . Cambillau, C. (2016). The Atomic Structure of the Phage Tuc2009 Baseplate Tripod Suggests that Host Recognition Involves Two Different Carbohydrate Binding Modules. *mBio, 7*(1), e01781.

Leiman, P. G., Arisaka, F., van Raaij, M. J., Kostyuchenko, V. A., Aksyuk, A. A., Kanamaru, S., & Rossmann, M. G. (2010). Morphogenesis of the T4 tail and tail fibers. *Virology journal, 7*, 355-355.

Leiman, P. G., Chipman, P. R., Kostyuchenko, V. A., Mesyanzhinov, V. V., & Rossmann, M. G. (2004). Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell, 118*(4), 419-429.

Leiman, P. G., & Shneider, M. M. (2012). Contractile Tail Machines of Bacteriophages. In (2012 ed., Vol. 726, pp. 93-114). Boston, MA: Springer US.

Leonard, K. R., Kleinschmidt, A. K., & Lake, J. A. (1973). Caulobacter crescentus bacteriophage phiCbK: structure and in vitro self-assembly of the tail. *J Mol Biol, 81*(3), 349-365.

Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., . . . Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature methods, 10*(6), 584-590.

Liu, C., Perilla, J. R., Ning, J., Lu, M., Hou, G., Ramalho, R., . . . Zhang, P. (2016). Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nature communications, 7*, 10714-10714.

Liu, J., Yong, W., Deng, Y., Kallenbach, N. R., & Lu, M. (2004). Atomic structure of a tryptophan-zipper pentamer. *Proceedings of the National Academy of Sciences of the United States of America, 101*(46), 16156.

Luria, S. E., Delbrück, M., & Anderson, T. F. (1943). Electron Microscope Studies of Bacterial Viruses. *Journal of bacteriology, 46*(1), 57-77.

Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology, 152*(1), 36-51.

Mc Grath, S., Neve, H., Seegers, J. F. M. L., Eijlander, R., Vegge, C. S., Brøndsted, L., . . . van Sinderen, D. (2006). Anatomy of a lactococcal phage tail. *Journal of bacteriology, 188*(11), 3972-3982.

McHugh, C. A., Fontana, J., Nemecek, D., Cheng, N., Aksyuk, A. A., Heymann, J. B., . . . Hoiczyk, E. (2014). A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *The EMBO Journal, 33*(17), 1896-1911.

McMullan, G., Chen, S., Henderson, R., & Faruqi, A. R. (2009). Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy, 109*(9), 1126-1143.

Meier, B. H., Riek, R., & Bockmann, A. (2017). Emerging Structural Understanding of Amyloid Fibrils by Solid-State NMR. *Trends Biochem Sci, 42*(10), 777-787.

Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C., & Ferrin, T. E. (2006). Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC bioinformatics, 7*, 339-339.

Mindell, J. A., & Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol, 142*(3), 334-347.

Montag, D., & Henning, U. (1987). An open reading frame in the Escherichia coli bacteriophage lambda genome encodes a protein that functions in assembly of the long tail fibers of bacteriophage T4. *Journal of bacteriology, 169*(12), 5884-5886.

Montag, D., Schwarz, H., & Henning, U. (1989). A component of the side tail fiber of Escherichia coli bacteriophage lambda can functionally replace the receptor-recognizing part of a long tail fiber protein of the unrelated bacteriophage T4. *Journal of bacteriology, 171*(8), 4378-4384.

Moody, M. F. (1999). Geometry of phage head construction. *Journal of Molecular Biology, 293*(2), 401-433.

Mooney, P., Contarato, D., Denes, P., Gubbens, A., Lee, B., Lent, M., & Agard, D. (2011). A High-Speed Electron-Counting Direct Detection Camera for TEM. *Microscopy and Microanalysis, 17*(S2), 1004-1005.

Morais, M. C., Choi, K. H., Koti, J. S., Chipman, P. R., Anderson, D. L., & Rossmann, M. G. (2005). Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29. *Mol Cell, 18*(2), 149-159.

Mougous, J. D., Cuff, M. E., Raunser, S., Shen, A., Zhou, M., Gifford, C. A., . . . Mekalanos, J. J. (2006). A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus. *Science (New York, N.Y.), 312*(5779), 1526-1530.

Murialdo, H., & Siminovitch, L. (1972). The morphogenesis of bacteriophage lambda: IV. Identification of gene products and control of the expression of the morphogenetic information. *Virology, 48*(3), 785-823.

Murray, N. E., & Gann, A. (2007). What has phage lambda ever done for us? *Current Biology, 17*(9), R305-R312.

Nicastro, J., Sheldon, K., & Slavcev, R. A. (2014). Bacteriophage lambda display systems: developments and applications. *Appl Microbiol Biotechnol, 98*(7), 2853-2866.

Oh, B. (2015). *Multiple functional roles of the bacteriophage HK97 delta domain in capsid assembly.* (PhD Doctor of Philosophy Doctoral Dissertation), Univerity of Pittsburgh,

Oh, B., Moyer, C. L., Hendrix, R. W., & Duda, R. L. (2014). The delta domain of the HK97 major capsid protein is essential for assembly. *Virology, 456-457*, 171-178.

Pell, L. (2010). *The structure, evolution, and assembly mechanism of the bacteriophage tail tube.* (Ph.D.), University of Toronto,

Pell, L. G., Gasmi-Seabrook, G. M. C., Morais, M., Neudecker, P., Kanelis, V., Bona, D., . . . Maxwell, K. L. (2010). The Solution Structure of the C-Terminal Ig-like Domain of the Bacteriophage λ Tail Tube Protein. *Journal of Molecular Biology, 403*(3), 468-479.

Pell, L. G., Kanelis, V., Donaldson, L. W., Howell, P. L., & Davidson, A. R. (2009). The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proceedings of the National Academy of Sciences of the United States of America, 106*(11), 4160-4165.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem, 25*(13), 1605-1612.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera–A visualization system for exploratory research and analysis. *J Comput Chem, 25*(13), 1605-1612.

Pietilä, M. K., Laurinmäki, P., Russell, D. A., Ko, C.-C., Jacobs-Sera, D., Hendrix, R. W., . . . Butcher, S. J. (2013). Structure of the archaeal head-tailed virus HSTV-1 completes the HK97 fold story. *Proceedings of the National Academy of Sciences, 110*(26), 10604.

Plisson, C., White, H. E., São-José, C., Tavares, P., Zafarani, A., Auzat, I., . . . Lhuillier, S. (2007). Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. *The EMBO Journal, 26*(15), 3720-3728.

Popa, M. P., McKelvey, T. A., Hempel, J., & Hendrix, R. W. (1991). Bacteriophage HK97 structure: wholesale covalent cross-linking between the major head shell subunits. *Journal of virology, 65*(6), 3227-3237.

Rajagopala, S. V., Casjens, S., & Uetz, P. (2011). The protein interaction map of bacteriophage lambda. *BMC Microbiology, 11*(1), 213.

Rao, V. B., & Black, L. W. (2010). Structure and assembly of bacteriophage T4 head. *Virology journal, 7*(1), 356-356.

Rigden, D. J., Jedrzejas, M. J., & Galperin, M. Y. (2003). Amidase domains from bacterial and phage autolysins define a family of γ-d,l-glutamate-specific amidohydrolases. *Trends in Biochemical Sciences, 28*(5), 230-234.

Roessner, C. A., & Ihler, G. M. (1984). Proteinase sensitivity of bacteriophage lambda tail proteins gpJ and pH in complexes with the lambda receptor. *Journal of bacteriology, 157*(1), 165-170.

Ross, P. D., Cheng, N., Conway, J. F., Firek, B. A., Hendrix, R. W., Duda, R. L., & Steven, A. C. (2005). Crosslinking renders bacteriophage HK97 capsid maturation irreversible and effects an essential stabilization. *The EMBO Journal, 24*(7), 1352-1363.

Ruska, H. (1940). Die Sichtbarmachung der bakteriophagen Lyse im Übermikroskop. *Naturwissenschaften, 28*(3), 45-46.

Scandella, D., & Arber, W. (1976). Phage lambda DNA injection into Escherichia coli pel- mutants is restored by mutations in phage genes V or H. *Virology, 69*(1), 206-215.

Scheres, S. H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol, 180*(3), 519-530.

Sciara, G., Bebeacua, C., Bron, P., Tremblay, D., Ortiz-Lombardia, M., Lichière, J., . . . Cambillau, C. (2010). Structure of lactococcal phage p2 baseplate and its mechanism of activation. *Proceedings of the National Academy of Sciences of the United States of America, 107*(15), 6852-6857.

Searle, B. C. (2010). Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics, 10*(6), 1265-1269.

Sorzano, C. O. S., Alcorlo, M., de la Rosa-Trevín, J. M., Melero, R., Foche, I., Zaldívar-Peraza, A., . . . Carazo, J. M. (2015). Cryo-EM and the elucidation of new macromolecular structures: Random Conical Tilt revisited. *Scientific reports, 5*, 14290.

Spinelli, S., Veesler, D., Bebeacua, C., & Cambillau, C. (2014). Structures and host-adhesion mechanisms of lactococcal siphophages. *Frontiers in microbiology, 5*, 3-3.

Stone, J. E., Gullingsrud, J., & Schulten, K. (2001). *A system for interactive molecular dynamics simulation*. Paper presented at the ACM Symposium on Interactive 3D graphics, New York.

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification, 41*(1), 207-234.

Suhanovsky, M. M., & Teschke, C. M. (2015). Nature's favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. *Virology, 479-480*, 487-497.

Suiter, C. L., Quinn, C. M., Lu, M., Hou, G., Zhang, H., & Polenova, T. (2015). MAS NMR of HIV-1 protein assemblies. *Journal of magnetic resonance (San Diego, Calif. : 1997), 253*, 10-22.

Sutter, M., Boehringer, D., Gutmann, S., Gunther, S., Prangishvili, D., Loessner, M. J., . . . Ban, N. (2008). Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat Struct Mol Biol, 15*(9), 939-947.

Tam, W., Pell, L. G., Bona, D., Tsai, A., Dai, X. X., Edwards, A. M., . . . Davidson, A. R. (2013). Tail tip proteins related to bacteriophage λ gpL coordinate an iron-sulfur cluster. *Journal of Molecular Biology, 425*(14), 2450-2462.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., & Ludtke, S. J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol, 157*(1), 38-46.

Taylor, N. M., Prokhorov, N. S., Guerrero-Ferreira, R. C., Shneider, M. M., Browning, C., Goldie, K. N., . . . Leiman, P. G. (2016). Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature, 533*(7603), 346-352.

Tetz, G., Brown, S. M., Hao, Y., & Tetz, V. (2018). Parkinson's disease and bacteriophages as its overlooked contributors. *Scientific reports, 8*(1), 10812.

Thomas, J. O., Sternberg, N., & Weisberg, R. (1978). Altered arrangement of the DNA in injection-defective lambda bacteriophage. *Journal of Molecular Biology, 123*(2), 149-161.

Tivol, W. F., Briegel, A., & Jensen, G. J. (2008). An improved cryogen for plunge freezing. *Microsc Microanal, 14*(5), 375-379.

Trabuco, L. G., Villa, E., Mitra, K., Frank, J., & Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure (London, England : 1993), 16*(5), 673-683.

Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B., & Schulten, K. (2009). Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods (San Diego, Calif.), 49*(2), 174-180.

Tso, D.-j., Hendrix, R. W., & Duda, R. L. (2014). Transient contacts on the exterior of the HK97 procapsid that are essential for capsid assembly. *Journal of Molecular Biology, 426*(10), 2112-2129.

Tsugita, A., Black, L. W., & Showe, M. K. (1975). Protein cleavage during virus assembly: characterization of cleavage in T4 phage. *J Mol Biol, 98*(1), 217-215.

Tsui, L., & Hendrix, R. W. (1980). Head-tail connector of bacteriophage lambda. *Journal of Molecular Biology, 142*(3), 419-438.

Tsui, L. C., & Hendrix, R. W. (1983). Proteolytic processing of phage lambda tail protein gpH: timing of the cleavage. *Virology, 125*(2), 257-264.

Twort, F. W. (1925). THE DISCOVERY OF THE "BACTERIOPHAGE.". *The Lancet, 205*(5303), 845.

Veesler, D., Khayat, R., Krishnamurthy, S., Snijder, J., Huang, R. K., Heck, A. J. R., . . . Johnson, J. E. (2014). Architecture of a dsDNA viral capsid in complex with its maturation protease. *Structure (London, England : 1993), 22*(2), 230-237.

Veesler, D., Robin, G., Lichière, J., Auzat, I., Tavares, P., Bron, P., . . . Cambillau, C. (2010). Crystal structure of bacteriophage SPP1 distal tail protein (gp19.1): a baseplate hub paradigm in gram-positive infecting phages. *The Journal of biological chemistry, 285*(47), 36666-36673.

Vernhes, E., Renouard, M., Gilquin, B., Cuniasse, P., Durand, D., England, P., . . . Boulanger, P. (2017). High affinity anchoring of the decoration protein pb10 onto the bacteriophage T5 capsid. *Scientific reports, 7*, 41662-41662.

Vianelli, A., Wang, G. R., Gingery, M., Duda, R. L., Eiserling, F. A., & Goldberg, E. B. (2000). Bacteriophage T4 self-assembly: localization of gp3 and its role in determining tail length. *Journal of bacteriology, 182*(3), 680-688.

Walker, J. E., Auffret, A. D., Carne, A., Gurnett, A., Hanisch, P., Hill, D., & Saraste, M. (1982). Solid-Phase Sequence Analysis of Polypeptides Eluted from Polyacrylamide Gels. *European Journal of Biochemistry, 123*(2), 253-260.

Wang, J., Hofnung, M., & Charbit, A. (2000). The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of Escherichia coli K-12. *Journal of bacteriology, 182*(2), 508-512.

Ward, M. E., Ritz, E., Ahmed, M. A. M., Bamm, V. V., Harauz, G., Brown, L. S., & Ladizhansky, V. (2015). Proton detection for signal enhancement in solid-state NMR experiments on mobile species in membrane proteins. *J Biomol NMR, 63*(4), 375-388.

Weigle, J. (1966). Assembly of phage lambda in vitro. *Proceedings of the National Academy of Sciences of the United States of America, 55*(6), 1462-1466.

Wikoff, W. R., Che, Z., Duda, R. L., Hendrix, R. W., & Johnson, J. E. (2003). Crystallization and preliminary analysis of a dsDNA bacteriophage capsid intermediate: Prohead II of HK97. *Acta Crystallogr D Biol Crystallogr, 59*(Pt 12), 2060-2064.

Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W., & Johnson, J. E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. *Science, 289*(5487), 2129-2133.

William Studier, F., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990). [6] Use of T7 RNA polymerase to direct expression of cloned genes. In *Methods in Enzymology* (Vol. 185, pp. 60-89): Academic Press.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., . . . Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr, 67*(Pt 4), 235-242.

Xie, Z., & Hendrix, R. W. (1995). Assembly in vitro of bacteriophage HK97 proheads. *J Mol Biol, 253*(1), 74-85.

Xu, J. (2001). *A CONSERVED FRAMESHIFT STRATEGY IN dsDNA LONG TAILED BACTERIOPHAGES.* (Doctor of Philosophy), University of Pittsburgh,

Xu, J., Hendrix, R. W., & Duda, R. L. (2013). A balanced ratio of proteins from gene G and frameshift-extended gene GT is required for phage lambda tail assembly. *Journal of Molecular Biology, 425*(18), 3476-3487.

Xu, J., Hendrix, R. W., & Duda, R. L. (2014). Chaperone–Protein Interactions That Mediate Assembly of the Bacteriophage Lambda Tail to the Correct Length. *Journal of Molecular Biology, 426*(5), 1004-1018.

Yan, X., Dryden, K. A., Tang, J., & Baker, T. S. (2007). Ab initio random model method facilitates 3D reconstruction of icosahedral particles. *J Struct Biol, 157*(1), 211-225.

Yang, F., Forrer, P., Dauter, Z., Conway, J. F., Cheng, N., Cerritelli, M. E., . . . Wlodawer, A. (2000). Novel fold and capsid-binding properties of the λ-phage display platform protein gpD. *Nature Structural Biology, 7*, 230.

Yap, M. L., Klose, T., Arisaka, F., Speir, J. A., Veesler, D., Fokine, A., & Rossmann, M. G. (2016). Role of bacteriophage T4 baseplate in regulating assembly and infection. *Proceedings of the National Academy of Sciences of the United States of America, 113*(10), 2654-2659.

Youderian, P. A. (1978). (Doctor of Philosophy Ph.D.), Massachusetts Institute of Technology, Cambridge, Massachusetts.

Zaccai, N. R., Chi, B., Thomson, A. R., Boyle, A. L., Bartlett, G. J., Bruning, M., . . . Woolfson, D. N. (2011). A de novo peptide hexamer with a mutable channel. *Nature Chemical Biology, 7*, 935.

Zairi, M., Stiege, A. C., Nhiri, N., Jacquet, E., & Tavares, P. (2014). The collagen-like protein gp12 is a temperature-dependent reversible binder of SPP1 viral capsids. *The Journal of biological chemistry, 289*(39), 27169-27181.

Zeng, M., Shang, Y., Araki, Y., Guo, T., Huganir, R. L., & Zhang, M. (2016). Phase Transition in Postsynaptic Densities Underlies Formation of Synaptic Complexes and Synaptic Plasticity. *Cell, 166*(5), 1163-1175.e1112.

Zheng, W., Wang, F., Taylor, N. M. I., Guerrero-Ferreira, R. C., Leiman, P. G., & Egelman, E. H. (2017). Refined Cryo-EM Structure of the T4 Tail Tube: Exploring the Lowest Dose Limit. *Structure (London, England : 1993), 25*(9), 1436-1441.e1432.

Zinke, M., Fricke, P., Samson, C., Hwang, S., Wall, J. S., Lange, S., . . . Lange, A. (2017). Bacteriophage Tail-Tube Assembly Studied by Proton-Detected 4D Solid-State NMR. *Angewandte Chemie (International ed. in English), 56*(32), 9497-9501.

Zivanovic, Y., Confalonieri, F., Ponchon, L., Lurz, R., Chami, M., Flayhan, A., . . . Boulanger, P. (2014). Insights into bacteriophage T5 structure from analysis of its morphogenesis genes and protein components. *Journal of virology, 88*(2), 1162-1174.